IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Exploration of Opinion Mining with Machine Learning

S.Vedantharajagopal

Dr. S. Babu

Abstract: Opinion mining, also known as sentiment analysis, is a crucial task in natural language processing that aims to extract and understand subjective information from text data. With the proliferation of online content, social media, and customer reviews, analyzing and summarizing opinions have become increasingly important for businesses, governments, and researchers. Machine learning techniques have played a pivotal role in advancing the field of opinion mining, enabling automated sentiment classification, aspect-based analysis, and trend prediction.Machine learning algorithms, especially supervised learning techniques, can be trained to automatically classify text into sentiment categories such as positive, negative, or neutral. These algorithms learn from labeled data, making them capable of processing vast amounts of text and providing sentiment scores at scale.Machine learning models can go beyond simple positive/negative classification to perform fine-grained sentiment analysis. They can classify sentiment on a scale, detecting sentiments like strongly positive, mildly positive, strongly negative, mildly negative, or neutral. This work is indented to interpret most recent novel text data opinion mining methodologies, their advantages and the limitations.

Keywords: Opinion Mining, Sentiment Analysis, Machine Learning, Automatic sentiment classification, Contextual screening

1. Introduction

Online stores like Amazon rely heavily on customer feedback and product reviews. Opinion mining helps in automatically analyzing and summarizing the sentiments expressed in these reviews. It allows Amazon to quickly understand how customers perceive their products and services.By analyzing customer opinions, Amazon can identify areas for product improvement. For example, if customers consistently express negative sentiments about a particular feature or aspect of a product, Amazon can use this feedback to make necessary adjustments and enhancements [1].Opinion mining provides valuable insights into market trends and consumer preferences. Amazon can analyze sentiments to identify emerging trends, popular product categories, and changes in customer preferences, allowing them to make informed decisions about inventory, marketing, and product development [2].

Maintaining a positive online reputation is crucial for e-commerce platforms. Opinion mining helps Amazon monitor its brand reputation by identifying and addressing negative sentiments or complaints promptly. This can lead to improved customer satisfaction and trust.Amazon can use sentiment analysis to gain insights into how their products and services compare to those of competitors [3]. By analyzing customer reviews and opinions, they can identify areas where they excel and areas where they may need to catch up or differentiate themselves.Opinion mining can be applied to customer service interactions, including chat logs and customer support emails. Analyzing sentiments in these interactions can help Amazon identify common pain points and improve customer service processes [4].Understanding customer sentiments allows Amazon to provide personalized recommendations and product suggestions. By analyzing past purchase history and sentiment data, Amazon can offer products and content that align with individual customer preferences.

Sentiment analysis can inform Amazon's marketing strategies. By understanding what aspects of products or services customers value most, they can tailor their marketing messages and advertising campaigns to resonate with target audiences.Opinion mining can be used to identify counterfeit or low-quality products by analyzing customer feedback. Amazon can take corrective actions to remove or verify such products, enhancing trust in their marketplace.Text-based opinion mining involves various machine learning algorithms and techniques that can be applied depending on the specific task and requirements. Here are some commonly used machine learning algorithms and approaches for text-based opinion mining:

Naive Bayes classifiers are simple and efficient for sentiment analysis tasks [5]. They are often used for binary sentiment classification (positive/negative). They rely on Bayes' theorem and assume independence between words.Logistic regression is another common choice for binary sentiment classification. It models the probability of a given text belonging to a particular sentiment class. SVMs are powerful classifiers used for both binary and multi-class sentiment classification [6]. They work by finding the hyperplane that best separates data points of different sentiment classes.Decision tree algorithms, such as Random Forests and Gradient Boosted Trees, are used for sentiment analysis tasks, especially when interpretability is important. They can handle both binary and multi-class sentiment classification.Deep learning approaches, including feedforward neural networks and recurrent neural networks (RNNs) [7], have become increasingly popular for sentiment analysis tasks. Long Short-Term Memory (LSTM) [8] and Gated Recurrent Unit (GRU) [9] networks are commonly used for sequence data like text.

Transformers, including models like BERT, GPT, and RoBERTa, have revolutionized sentiment analysis by capturing complex contextual information in text. Pre-trained transformer models are fine-tuned for various sentiment analysis tasks, achieving state-of-the-art results.Ensemble techniques like Bagging and Boosting can be used to combine multiple base models (e.g., decision trees, SVMs, or neural networks) to improve the overall performance of sentiment analysis systems.

Word embeddings like Word2Vec, GloVe, and FastText are used to represent words in a continuous vector space. These embeddings can enhance the features used by other machine learning algorithms, providing richer contextual information [10].Sequence-to-sequence models, including variants like Seq2Seq with Attention, can be used for tasks like text generation and summarization, which are related to opinion mining.Reinforcement learning can be applied to tasks like sentiment-based recommendation systems. Agents can be trained to make recommendations based on user sentiment and feedback.

There are some notable extensions and novel approaches over the traditional machine learning algorithms are attempted to get better results. A clear investigation is carried out in this work to understand the basic principles of such works with their merits and limitations to conclude about the research gap and objectives to be carried over towards the advancements.

2. Existing methods

There are several existing works formulated by many researchers for the purpose of using machine learning for opinion mining from text data. A set of most recent state-of-the-art works are undertaken here to study the strategies. This section elaborates the methodologies, advantages and limitations of the selected works in detail.

2.1. A Novel Approach for Sentiment Analysis and Opinion Mining on Social Media Tweets

In 2023, Nagesh Yagnam introduced this work [11] to apply machine learning in opinion mining. The primary objective of this work is to delve into the sentiments and opinions of content creators about various subjects, as well as the comments of social media users within the context of Ethiopia, focusing on the Facebook platform. To achieve this, the paper employs natural language processing (NLP) and machine learning (ML) techniques within a novel sentiment analysis and opinion mining framework. The goal is to gain insights from the posts and user comments on social media, specifically within the realm of Facebook. The idea of the utilization of sentiment-specific embeddings, referred to as "sentiment embeddings," as a novel approach to enhance sentiment analysis is the core of the work. To bolster the quality of continuous word representations, the author leverages the sentiment conveyed in the text to maintain the relevance and potency of word contexts. Word level sentiment analysis, Queueing of sentiment related keywords, and Sentence level sentiment classification are focused with fresh approach in this work.

Application of ML in opinion mining saves much human efforts and time, which are noted as the advantages of this work. The classification requires more accuracy these days to be on par with the competitive industries. Realization of only moderate accuracy is known as the limitation of this work.

2.2. Multi-lingual opinion mining for social media discourses: an approach using deep learning based hybrid fine-tuned smith algorithm with Adam optimizer (MOMDL)

In 2023, Aniket K. Shahade et.al., commenced MOMDL work [12] work for opinion mining using deep learning methodology. A novel combination of techniquesintroduced in MOMDLwork for conducting multilingual sentiment analysis through a deep learning approach, which combines a refined Smith algorithm with the Adam optimizer, proves to be highly efficient in extracting valuable information from social media content across various languages. This method has showcased exceptional levels of accuracy, precision, recall, and F1-Score metrics, establishing its utility as a valuable tool for enhancing performance and customer contentment by uncovering patterns and trends in public sentiment.Furthermore, the suggested approach surpasses alternative existing techniques, such as PGM, MCM, CNN, and NBi-LSTM, not only in computational efficiency but also in terms of overall performance. This demonstrates its substantial contribution to the realm of information management. The preprocessing stage in MOMDL work contains several steps such as Tokenization, Parts of speech tagging, Stop-word removal, Stemming and Lemmatization. Text vectorization process is carried over through Naïve Bayes vectorization, Laplace smoothing method, and Enhanced Naïve Bayes vectorization in MOMDL work.

Achievement of higher classification accuracy is the stated advantage of MOMDL work. The multi-state preprocessing procedure causes a sensible delay in the processing. The increased processing time may not be suitable for real-time streaming text data processing, which is identified as the limitation of MOMDL work.

2.3. Artificial fish swarm optimization with deep learning enabled opinion mining approach (AFSODL)

Saud S.Alotaibi et.al., introduced AFSODL work [13] in 2022 for the purpose of applying optimized deep learning methodology for the purpose of opinion mining from text data. A novel Artificial Fish Swarm Optimization with Bidirectional Long Short Term Memory (AFSO-BLSTM) model has been created for Opinion Mining (OM). The primary objective of the AFSO-BLSTM model is to efficiently extract and analyze opinions embedded within textual data. Furthermore, the AFSO-BLSTM model undergoes preprocessing and employs TF-IDF-based feature extraction. Additionally, it employs the BLSTM model for effective opinion detection and classification. Notably, the AFSO algorithm is harnessed to fine-tune the hyperparameters of the BLSTM model, highlighting the originality of this work. The comprehensive simulation study of the AFSO-BLSTM model is validated using a benchmark dataset, and the experimental results underscore the substantial potential of the AFSO-BLSTM model in opinion mining. The experiments are performed for the IMDB dataset.

Higher Accuracy and Precision towards opinion mining are achieved by AFSODL work, that stands by the advantage side, whereas fluctuating convergence during training process shows the instability of AFSODL work, which is observed as the limitation

2.4. A new opinion mining method based on M-cuckoo with M-cat and SVM with NB classification algorithm(SVMNB)

Malathi M et.al., submitted SVMNB work [14] in 2022 in the motivation of bringing up a new ensemble that combines M-Cat, M-Cuckoo with SVM and Naïve Mayes classification algorithms. SVMNB seeks to discern the writer's or speaker's emotional disposition or stance towards the product, which can range from positive to negative. Human emotions, whether positive or negative, are commonly referred to as sentiments or feelings. Opinion mining, frequently referred to as sentiment analysis, is a branch of computational linguistics and natural language processing (NLP) that specializes in identifying and extracting subjective content from source materials [1]. Given the escalating popularity of e-commerce, which serves as a common avenue for the expression and exchange of thoughts, emotions, and ideas, evaluations and reviews are becoming increasingly conspicuous. The experiments are performed for Amazon appliances dataset.

Application of multiple optimization techniques and classification algorithm ensembles is stated as the novelty of this work. Shorter processing time due to multiple optimizations is the advantage of SVMNB work.At the same time, the performance scores subject to Accuracy, Precision, and F-Score are up to the mark, which is understood as the limitation of SVMNB work.

2.5. Multus Medium Opinion Mining (MMOM) A Novel Hybrid Deep Learning Approach for Comprehensive Product Review Analysis

In 2023, MMOM [15] work is submitted by Bushra Kanwal et.al., for the purpose of employing novel hybrid deep learning method for the purpose of comprehensive opinion mining. MMOM work presents a groundbreaking deep learning methodology called Multus Medium Opinion Mining (MMOM), which leverages both text and image data to conduct comprehensive analysis of product reviews. MMOM employs an integrated model that combines Bidirectional Long Short-Term Memory (BiLSTM) and embedded Convolutional Neural Network (CNN) architectures, integrating elements from GoogleNet and VGGNet. This approach facilitates the efficient extraction and merging of textual and visual features. This collaborative strategy encompasses data collection, preprocessing, feature extraction, the generation of feature vectors based on fusion strategies, and subsequent product recommendations. Performance assessment was conducted using two diverse real-world datasets, namely "flicker8k" and "t4sa," revealing substantial enhancements compared to existing methodologies.Feature extraction using GoogleNet and VGGNet, Deep Feature Extraction, Execution of Fusion and the application of Reinforcement Learning are the different stated used in MMOM work.

Achievement of higher Accuracy and Precision is the noted advantage of MMOM work. Dependency on multi-stage processing increases the processing time significantly. Thus, application of MMOM work increases the computational and time complexity, which is noted as the limitation.

2.6. Multi-tier sentiment analysis of social media text using supervised machine learning (MSASML)

In 2022, MSASML work [16] is proposed by Hameedur Rahman et.al., to test the opinion mining ability of different machine learning algorithms along with multi-tier architectures. An exploration of pre-processing techniques and machine learning models for conducting multi-class sentiment classification is performed in MSASML work. To enhance performance, a multi-layer classification model is introduced in MSASMLThe authors implemented supervised machine learning models, specifically Decision Tree, Support Vector Machine, and Naïve Bayes, for the task of sentiment classification. MSASML compares the performance of single-layer architecture models with the multi-tier model. The results indicate a slight improvement in the multi-tier model's performance compared to the single-layer architecture. Additionally, the multi-tier models exhibit better recall, enabling the proposed model to capture more contextual information.

Lower computational complexity is identified as the advantage, whereas only moderate performance in scoring accuracy and precision parameters is the observed limitation of MSASML work

2.7. A Machine Learning-Based Technique with IntelligentWordNet Lemmatize for Twitter Sentiment Analysis(TF-IDF-RF)

In 2022, S.Saranya et.al., introduced a work [17] for opinion mining through machine learning using TF-TDF concept. In TF-IDF-RF work, the authors proposed a machine-learning-based approach for sentiment analysis that navigates these complexities. TF-IDF-RF method involves the extraction of features using Term Frequency-Inverse Document Frequency (TF-IDF) and incorporates deep intelligent wordnet lemmatization to enhance the quality of tweets by filtering out noise and standardizing language usage. To detect the emotional content of a tweet, the authors employ the Random Forest algorithm, a robust tool in machine learning. To validate the efficacy of TF-IDF-RF proposed approach, it is tested rigorouslyusing publicly available datasets. The results of TF-IDF-RF experiments demonstrate a significant improvement in sentiment classification, particularly in the context of multi-class emotional text data. Each phase of TF-IDF-RF work is carried out using different principles such as, Pre-processing is done by WordNet Lemmatize, Feature extraction is performed by TF-IDF, Dimension reduction is carried out using Latent Direct Allocation method, and the classification is handled by Random Forest method.

On the whole the experiments carried out in TF-IDF-RF work shows that the attained accuracy is 93%, which is the advantage of this work. The TF-IDF based feature extraction makes the convergence process delay with term frequency fluctuations. These fluctuations act as a barrier in attaining higher accuracy, which is identified as the limitation of TF-IDF-RF work.

2.8. An enhanced approach for sentiment analysis based on meta-ensemble deep learning (SAMEDL)

In 2023, Rania Kora et.al., presented a work [18] for applying meta-ensemble deep learning method for the purpose of sentiment analysis of text data. A meta-ensemble deep learning approach is proposed in SAMEDL work to enhance sentiment analysis performance. In this approach, baseline deep learning models are trained and fused using three levels of meta-learners. Then a benchmark dataset called "Arabic-Egyptian Corpus 2" is introduced as an extension of a previous corpus. The corpus size has been expanded by 10,000 annotated tweets written in colloquial Arabic on various topics. Afterward several experiments are conducted on six benchmark datasets for sentiment analysis in different languages and dialects to assess the effectiveness of the proposed meta-ensemble deep learning approach. The experimental results demonstrate that the meta-ensemble approach consistently outperforms the baseline deep learning models. Additionally, the experiments indicate that meta-learning further enhances performance when using probability class distributions to train the meta-learners.

Achievement of higher accuracy by SAMEDL ensembles than the other compared methods is the advantage of this work. More optimizations are required to get some more accuracy and precision which is known as the limitation of SAMEDL work.

2.9. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis (DLMHFE)

In 2023, DLMHFE work [19] is put forward by Gagandeep Kaur et.al., A hybrid approach for analyzing sentiments is presented in DLMHFE work. The process involves pre-processing, feature extraction, and sentiment classification. Utilizing NLP techniques, the pre-processing stage eliminates undesirable data from input text reviews. To extract features effectively, a hybrid method combining review-related features and aspect-related features is introduced to construct the distinctive hybrid feature vector corresponding to each review. Sentiment classification is performed using the deep learning classifier LSTM. DLMHFE procedure was experimentally evaluated using three different research datasets.

Suggesting a novel technique for sentiment analysis, with a special focus on consumer review summarization, based on the aforementioned concerns of state-of-the-art approaches. The suggested method consists of three stages: pre-processing, feature extraction, and sentiment classification. To denoise input reviews, the pre-processing step employs a standard procedure. The feature extraction step is ingeniously designed as a hybrid step to address the challenges of accurately representing the features of input reviews. Utilization of Hybrid feature vector, Review related features, Aspect rotated features, and LSTM based sentiment classification altogether establishes the DLMHFE work.

Achievement of higher Accuracy, Precision and F-Score are the observed advantages of DLMHFE method, whereas ensemble of multiple classification algorithms causes a notable increase in computational complexity. Due to the increased computational complexity, DLMHFE work may not be able to handle large datasets, which is identified as the limitation.

2.10. Convolutional neural network based sentiment analysis with tf-idf based vectorization (CNN-TFIDF)

Bhavna Kabra et.al., proposed CNN-TFIDF work [20] at 2023, for the purpose of applying CNN over the TF-IDF based vectorization to analyze sentiment of text input data. A novel approach to sentiment analysis utilizing deep learning is introduced, distinguishing it from established techniques that maintain a fixed sample size for input data while augmenting the proportion of sentiment-related data in each review. The study introduces models within the CNN-TFIDF family, which concurrently integrate CNN and TF-IDF techniques, drawing inspiration from the latest advancements in neural network research. Experiments conducted on challenging datasets demonstrate that the proposed methodology outperforms numerous conventional approaches. This model exhibits superior performance compared to traditional machine learning methods, achieving an accuracy rate of 87 percent. In contrast to prior endeavors, this study also achieves remarkably high levels of accuracy.

Initially, textual data is collected and prepared. Subsequently, data cleaning is conducted, an essential preprocessing step aimed at eliminating extraneous information, notations, and punctuations while condensing the text. For feature extraction, CNN-TFIDF uses TF-IDF, a technique that standardizes the representation of terms with similar meanings. This involves grouping words to construct the representation, and this process is facilitated by word embedding, which captures the relationships between words. In the context of sentiment analysis, this is commonly referred to as embedding. Multiple levels of embedding, including word, phrase, and document-level embedding, are performed. Finally, the document's polarity is assessed using a convolutional neural network as part of the sentiment analysis.

Integration of CNN and TF-IDF is the novelty of CNN-TFIDFwork. Accomplishment of moderate accuracy is stated as the advantage of this work. Incorporation of more optimization techniques is required to improve the accuracy score, which is understood as the limitation of CNN-TFIDF work.

2.11. PyFin-sentiment: Towards a machine-learning-based model for deriving sentiment from financial tweets (Py-Fin)

In 2023, Moritz Wilksch et.al., proposed an opinion mining method [21] for financial data. In Py-Finwork, a sentiment model is designed, trained, and deployed, demonstrating superior performance compared to all prior models (including VADER, NTUSD-Fin, FinBERT, TwitterRoBERTa) when assessed using Twitter posts. When applied to posts from a different platform, this model achieves performance comparable to BERT-based large language models. This outcome is attained while incurring significantly lower training and inference costs, thanks to the model's straightforward design. The resulting python library is made available to facilitate utilization by future researchers and practitioners. Data collection, Data sampling, Data labelling, and Data preprocessing are performed exclusively for this work by the authors.

Multiple machine learning models are trained on the cleaned data. A comparison is made between two machine learning models (logistic regression and a support vector machine) and three deep learning models (a recurrent neural network, a transformer neural network trained from scratch, and a BERT-based classification model). Experiments are conducted with both simple and complex models, with the simpler models selected for their speed and ability to establish a solid performance baseline. Recognizing that the majority of models in NLP are deep learning-based, the two most common architectures for text classification are introduced to the experiments and trained from scratch.PyTorch framework is used to implement the Py-Fin work.

Higher accuracy is the advantage of Py-Fin work, at the same time the inference time halts at 100mS denotes that there a complication in handling larger datasets with this model, which is the observed limitation.

2.12. Human behavior analysis on political retweets using machine learning algorithms (HBAMLA)

During the year of 2023, Het Patel et.al., proposed HBAMLA work [22] for the purpose sentiment analysis from text data by applying machine learning algorithms. The objective HBAMLAwork was to analyze tweets for political sentiments, specifically focusing on tweets related to prominent political leaders from various states and parties in India. The process involved is applying a polarity detection analysis to retweeted messages to develop a sentiment classification algorithm. The aim was to assess whether these tweets could predict the popularity of specific politicians among the general public. Comparisonof the subjectivity and polarity present in the tweets of these political leaders is also performed in HBAMLA work. Furthermore, the engagement levels of these leaders were considered as a factor in assessing their popularity. The findings of these comparisons were then visualized through data representations

Data extraction, Data preprocessing, and Data analysis are the different stages of HBAMLA work. Features such as Message length, likes, Replies, number of forwards, Quotes, Polarity and Subjectivity are used in HBAMLA for the purpose of sentiment analysis.

Attempt of applying machine learning algorithm for the purpose of opinion mining is found as the novelty of HBAMLA work. Benchmark parameters such as Accuracy, Precision and Sensitivity are not discussed, that is observed as the limitation of HBAMLA work.

2.13. Deep learning based sentiment analysis of public perception of working from home through tweets (DLBSA)

Aarushi Vohra et.al., introduced DLBSA [23] work during the year of 2022 for the purpose of opinion mining using CNN based deep learning procedure. A fine-tuned Convolutional Neural Network (CNN) model is proposed for the analysis of Twitter data. The deep learning model takes as input an annotated set of tweets, categorizing them into three sentiment classes: positive, negative, and neutral, using VADER (Valence Aware Dictionary for Sentiment Reasoning). In addition, FastText embeddings are utilized to modify the input vector to the embedding layer, aiding in the training of supervised word representations for a text corpus comprising over 450,000 tweets. The model employs multiple convolution and max pooling layers, dropout operations, and dense layers with ReLU and sigmoid activations, yielding impressive results on the dataset. Furthermore, the model's performance is compared to that of several standard classifiers, including Support Vector Machine (SVM), Naive Bayes, Decision Tree, and Random Forest.

Data Acquisition, Preprocessing, Data annotation, Data up sampling, Data segmentation, Tokenization, Padding, and Word embeddings are the phases taken care in DLBSA work. Achievement of 92.5% classification accuracy is the observed advantage of DLBSA work. Requirement of optimization techniques to improve the classification score is the identified limitation of DLBSA work.

2.14. Sentiment analysis using bidirectional LSTM network (SABLSTM)

During the year of 2023, U. B. Mahadevaswamy et.al., endeavored SABLSTM [24] work for performing opinion mining of text data using Bidirectional Long Short Term Memory (BiLSTM). SABLSTM work presents a technical overview of Sentiment Analysis employing a Bidirectional LSTM network. This model effectively manages prolonged dependencies by incorporating memory within the architecture to enhance prediction accuracy. The study utilizes the Amazon Product Review dataset, specifically examining 104,975 product reviews that capture users' sentiments regarding mobile electronic devices. The model's primary objective is to categorize the reviews into two groups: positive and negative sentiments. Ultimately, SABLSTM work summarizes the analysis findings and proposes potential directions for future research. A set of well-defined procedures are defined in SABLSTM work for the processes such as Data collection, Preprocessing, Word Encoding, Padding, Model design, Model training, and for Model evaluation,

The SABLSTM approach involving the Bidirectional LSTM model attains an elevated accuracy level of 91.4%. Subsequent research could involve a more detailed sentiment analysis, such as the categorization of sentiments into 3 to 5 gradations. The utilization of ensemble methods could potentially enhance the classifier's effectiveness. Moderate performance scores in terms of Accuracy, Precision, Recall, Specificity and F-Score is the observed advantage of SABLSTM work, More optimization techniques are required and hyperparameter tuning are required to improve the performance of SABLSTM which is known as the limitation.

2.15. MBi-GRUMCONV A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis

MBi-GRUMCONV work [25] has been proposed by Muhammet Sinan Başarslan et.al. during the year of 2023, for the purpose of testing the applicability of BiGRU and CNN ensemble in opinion mining of text data. MBi-GRUMCONV work introduced a novel deep learning model for sentiment analysis using the IMDB movie reviews dataset. The model conducts sentiment classification on vectorized reviews employing two Word2Vec methods, namely, Skip Gram and Continuous Bag of Words, across three distinct vector sizes (100, 200, 300). This is achieved through the integration of 6 Bidirectional Gated Recurrent Units and 2 Convolution layers (MBi-GRUMCONV). In the experiments performed with this model, the dataset was divided into training-test sets of 80%-20% and 70%-30%, with 10% of the training splits allocated for validation. Bidirectional recurrent neural networks utilize a unique architecture consisting of two independent Gated Recurrent Units (GRUs). This design guarantees that the networks constantly possess access to both preceding and subsequent information at each time step within the sequence.

The bidirectional GRU processes input data in two distinct manners: one from past to future and the other from future to past. What sets this approach apart from the unidirectional version is that the backward-working GRU retains future information and merges two hidden states. Consequently, the Bidirectional GRU has the capacity to retain information from both historical and future contexts at any given moment in the sequence. Furthermore, the ensemble of CNN along with BiGRU enables the method to gain more classification accuracy.

Classification performance was evaluated based on Accuracy and F-Sscore criteria. Notably, the proposed model achieved an accuracy rate of 95.34%, surpassing previous studies in the field. The experiments also indicated that Skip Gram contributed more significantly to the classification success which is the advantage of MBi-GRUMCONV work. Ensemble of two different deep learning methods one after another increase the processing time significantly, which is identified as the limitation of MBi-GRUMCONV work.

3. A crisp overview of the existing works, used methodologies, advantages and limitations are enumerated in Table 1.

Author	Work	Methodology	Advantages	Limitations
Nagesh Yagnam	A Novel Approach for Sentiment Analysis and Opinion Mining on Social Media Tweets	Sentiment Embeddings	Classification automation	Less Accuracy
Aniket K. Shahade et.al.,	Multi-lingual opinion mining for social media discourses: an approach using deep learning based hybrid fine-tuned smith algorithm with Adam optimizer	Nbi-LSTM	High Accuracy	High processing time
Saud S.Alotaibi et.al.,	Artificial fish swarm optimization with deep learning enabled opinion mining approach	Artificial Fish Swarm Optimization BiLSTM	Accuracy, Precision	Convergence Fluctuations
Malathi M et.al.,	A new opinion mining method based on M-cuckoo with M-cat and SVM with NB classification algorithm	M-Cat, M-Cuckoo and SVM	Less Processing Time	Low Accuracy Precision
Bushra Kanwal et.al.,	Multus Medium Opinion Mining (MMOM) A Novel Hybrid Deep Learning Approach for Comprehensive Product Review Analysis	BiLSTM CNN, VGGNet	High Accuracy	High computational complexity
Hameedur Rahman et.al.,	Multi-tier sentiment analysis of social media text using supervised machine learning	Decision Tree, SVM, Naïve Bayes	Less computaional complexity	Low Accuracy
S.Saranya et.al.,	A Machine Learning-Based Technique with IntelligentWordNet Lemmatize for Twitter Sentiment Analysis	TF-IDF Random Forest	Moderate Accuracy	Convergence Fluctuations
Rania Kora et.al.,	An enhanced approach for sentiment analysis based on meta-ensemble deep learning	Meta-Ensemble Deep Learning	High Accuracy	Missing Optimizations
Gagandeep Kaur et.al.,	A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis	NLP LSTM	High Accuracy, Precision	High computational complexity
Bhavna	Convolutional neural network based	CNN TF-IDF	Moderate	Missing

Kabra et.al.,	sentiment analysis with tf-idf based vectorization		Accuracy	Optimizations
Moritz Wilksch et.al.,	PyFin-sentiment: Towards a machine-learning-based model for deriving sentiment from financial tweets	VADER, NTUSD-Fin, FinBERT	Higher Accuracy	Datasize limitation
Het Patel et.al.	Human behavior analysis on political retweets using machine learning algorithms	Polarity Detection Analysis	Polarity detection	Missing Benchmark metrics
Aarushi Vohra et.al.,	Deep learning based sentiment analysis of public perception of working from home through tweets	CNN ReLU and SVM	Moderate Accuracy	Missing Optimizations
U.B.Mahadevaswamy et.al.,	Sentiment analysis using bidirectional LSTM network	BiLSTM	Accuracy, Precision, Recall	Missing Optimizations
Muhammet Sinan Başarslan et.al.	MBi-GRUMCONV A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis	Multi-BiGRU, CNN	High Accuracy, F-Score	High processing time

Table 1: Existing methods Key points

4. Scope for new Researches

Fine-grained Sentiment Analysis: Develop more advanced models for sentiment analysis that can categorize opinions into finer-grained sentiment classes beyond just positive, negative, and neutral, such as sentiment intensity, emotional analysis, or nuanced sentiments.Aspect-Based Sentiment Analysis: Extend research to focus on aspect-based sentiment analysis, where the goal is to identify sentiment toward specific aspects or features of a product or service mentioned in reviews.Multimodal Analysis: Combine text analysis with other modalities like images, audio, and video to provide a more comprehensive understanding of user sentiments and opinions. This is especially relevant for platforms like social media.Domain-specific Analysis: Explore domain-specific sentiment analysis for industries like healthcare, finance, or technology, where understanding user opinions can have significant implications.

Cross-Lingual Opinion Mining: Develop models and techniques for opinion mining across multiple languages, allowing businesses to analyze opinions from a global perspective.Temporal Analysis: Investigate how sentiment changes over time, considering factors such as seasonality, trends, and events. This can help in understanding evolving customer preferences and opinions.Bias and Fairness: Examine biases in opinion mining models and develop methods to mitigate them. Ensure fairness in sentiment analysis, especially in applications that involve decision-making.User Profiling: Create user profiles based on their opinions and sentiments, which can be used for personalized recommendations and targeted marketing.Summarization and Visualization: Develop tools for summarizing and visualizing large volumes of user reviews, making it easier for businesses to extract actionable insights.

5. Conclusion

There are several machine-learning based text data opinion mining procedures are in research and in practice. Though, it us realized that there are more requirements to carry over new research works in the field of opinion mining suing machine learning to coordinate with the developments in the field of Artificial Intelligence. Cross-domain Opinion Transfer and Real-time Opinion Mining are the critical need of time. Investigate methods for transferring knowledge learned from one domain of user reviews to another with limited labeled data, making it easier to adapt sentiment analysis models for new domains. Explore techniques for real-time opinion mining to enable businesses to respond quickly to emerging trends or issues.Investigate methods for transferring knowledge learned from one domain of user reviews to another with limited labeled data, making it easier to adapt sentiment analysis models for new domains.Explore techniques for real-time opinion mining to enable businesses to respond quickly to emerging trends or issues.

References:

[1] Al Hamli SS, Sobaih AEE. Factors Influencing Consumer Behavior towards Online Shopping in Saudi Arabia Amid COVID-19: Implications for E-Businesses Post Pandemic. Journal of Risk and Financial Management. 2023; 16(1):36. https://doi.org/10.3390/jrfm16010036

[2] Mustafa Kemal Yılmaz, Hale TuğçeAltunay, "Marketing insight from consumer reviews: Creating brand position through opinion mining approach," in Telematics and Informatics Reports, Volume 11, 2023, 100094, ISSN 2772-5030, https://doi.org/10.1016/j.teler.2023.100094

[3] Liu Y, Wan Y. Consumer Satisfaction with the Online Dispute Resolution on a Second-Hand Goods-Trading Platform. Sustainability. 2023; 15(4):3182. https://doi.org/10.3390/su15043182

[4] Abbas, Y., Malik, M.S.I. Defective products identification framework using online reviews. Electron Commer Res 23, 899–920 (2023). https://doi.org/10.1007/s10660-021-09495-8

[5] MiftahulQorib, Timothy Oladunni, Max Denis, Esther Ososanya, Paul Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," in Expert Systems with Applications, Volume 212, 2023, 118715, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.118715

[6] Quan, Z., Pu, L. An improved accurate classification method for online education resources based on support vector machine (SVM): Algorithm and experiment. Educ Inf Technol 28, 8097–8111 (2023). https://doi.org/10.1007/s10639-022-11514-6

[7] Chang hoon Song, Geonho Hwang , Jun ho Lee , Myungjoo Kang, "Minimal Width for Universal Property of Deep RNN," in Journal of Machine Learning Research 24 (2023), Link: https://www.jmlr.org/papers/volume24/22-1191/22-1191.pdf

[8] Gu Jianan, Ren Kehao, Gao Binwei, "Deep learning-based text knowledge classification for whole-process engineering consulting standards," in Journal of Engineering Research, 2023, ISSN 2307-1877, https://doi.org/10.1016/j.jer.2023.07.011

[9] Huang Y, Dai X, Yu J, Huang Z. SA-SGRU: Combining Improved Self-Attention and Skip-GRU for Text Classification. Applied Sciences. 2023; 13(3):1296. https://doi.org/10.3390/app13031296

[10] Avasthi, S., Chauhan, R. &Acharjya, D.P. Extracting information and inferences from a large text corpus. Int. j. inf. tecnol. 15, 435–445 (2023). https://doi.org/10.1007/s41870-022-01123-4

[11] Astarkie, M.G., Bala, B., Bharat Kumar, G.J., Gangone, S., Nagesh, Y. (2023). A Novel Approach for Sentiment Analysis and Opinion Mining on Social Media Tweets. In: Kumar, A., Ghinea, G., Merugu, S., Hashimoto, T. (eds) Proceedings of the International Conference on Cognitive and Intelligent Computing. Cognitive Science and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-19-2358-6_15

[12] Aniket K. Shahade, K.H. Walse, V.M. Thakare, Mohammad Atique, "Multi-lingual opinion mining for social media discourses: an approach using deep learning based hybrid fine-tuned smith algorithm with adam optimizer," in International Journal of Information Management Data Insights, Volume 3, Issue 2, 2023, 100182, ISSN 2667-0968, https://doi.org/10.1016/j.jjimei.2023.100182

[13] Saud S. Alotaibi, EatedalAlabdulkreem, Sami Althahabi, Manar Ahmed Hamza4, Mohammed Rizwanullah, Abu Sarwar Zamani, AbdelwahedMotwakel and Radwa Marzouk5, "Artificial Fish Swarm Optimization with Deep Learning Enabled Opinion Mining Approach," in Computer Systems Science & Engineering, Tech Science Press, https://doi.org/10.32604/csse.2023.030170

[14] Malathi, M., Thanamani, A. S., Padmapriya, P., Sharmila, S., Belwin, A. F., &Sherin, A. L. "A new opinion mining method based on M-cuckoo with M-cat and SVM with NB classification algorithm," in International Journal of Health Sciences, 6(S2), 12776–12785. https://doi.org/10.53730/ijhs.v6nS2.8353

[15] Kanwal, B.; Nawaz, A.; Mustafa, G.; Ali, T.; Babar, M.; Qureshi, B.; Koubaa, A. Multus Medium Opinion Mining (MMOM): A Novel Hybrid Deep Learning Approach for Comprehensive Product Review Analysis. Preprints 2023, 2023060405. https://doi.org/10.20944/preprints202306.0405.v1

[16] Bushra Kanwal, Asif Nawaz, Ghulam Mustafa, Tariq Ali, Muhammad Babar, Basit Qureshi, Anis Koubaa, "Multus Medium Opinion Mining (MMOM): A Novel Hybrid Deep Learning Approach for Comprehensive Product Review Analysis," in MDPI, Creative Commons CC BY, (www.preprints.org, https://doi.org/10.20944/preprints202306.0405.v1

[17] S. Saranya, G. Usha, "A Machine Learning-Based Technique with Intelligent WordNet Lemmatize for Twitter Sentiment Analysis," in Intelligent Automation & Soft Computing, Tech Science Press, https://doi.org/10.32604/iasc.2023.031987

[18] Kora, R., Mohammed, A. An enhanced approach for sentiment analysis based on meta-ensemble deep learning. *Soc. Netw. Anal. Min.* **13**, 38 (2023). https://doi.org/10.1007/s13278-023-01043-6

[19] Kaur, G., Sharma, A. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. J Big Data 10, 5 (2023). https://doi.org/10.1186/s40537-022-00680-6

[20] Kabra, B., & Nagar, C. (2023). Convolutional Neural Network based sentiment analysis with TF-IDF based vectorization. Journal of Integrated Science and Technology, 11(3), 503. Retrieved from https://pubs.thesciencein.org/journal/index.php/jist/article/view/503

[21] Moritz Wilksch, Olga Abramova, "PyFin-sentiment: Towards a machine-learning-based model for deriving sentiment from financial tweets," in International Journal of Information Management Data Insights, Volume 3, Issue 1, 2023, 100171, ISSN 2667-0968, https://doi.org/10.1016/j.jjimei.2023.100171

[22] Het Patel, Aditya Kansara, Boppuru Rudra Prathap, Kukatlapalli Pradeep Kumar, "Human behavior analysis on political retweets using machine learning algorithms," in Measurement: Sensors, Volume 27, 2023, 100768, ISSN 2665-9174, https://doi.org/10.1016/j.measen.2023.100768

[23] Vohra, A., Garg, R. Deep learning based sentiment analysis of public perception of working from home through tweets. J Intell Inf Syst 60, 255–274 (2023). https://doi.org/10.1007/s10844-022-00736-2

[24] U.B. Mahadevaswamy, P. Swathi, "Sentiment Analysis using Bidirectional LSTM Network," in Procedia Computer Science, Volume 218, 2023, Pages 45-56, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.12.400

[25] Başarslan, M.S., Kayaalp, F. MBi-GRUMCONV: A novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis. J Cloud Comp 12, 5 (2023). https://doi.org/10.1186/s13677-022-00386-3