Vol.67: e24240183, 2024
https://doi.org/10.1590/1678-4324-2024240183
ISSN 1678-4324 Online Edition

BABT
Brazilian Archives of Biology and Technology

Article - Engineering, Technology and Techniques

A Fused Feature Selection Technique for Enhanced Sentiment Analysis Using Deep Learning

Meenakshi Muthukrishnan [1*]
https://orcid.org/0000-0002-5181-5694

Suruliandi Andavar[1]
http://orcid.org/0000-0002-2863-1555

Raja Soosaimarian Peter Raj[2]
https://orcid.org/0000-0002-7216-2207

[1]ManonmaniamSundaranar University, Department of Computer Science and Engineering, Tirunelveli, Tamil Nadu, India; [2]Vellore Institute of Technology, School of Computer Science and Engineering, Vellore, Tamil Nadu, India.

Editor-in-Chief: Alexandre Rasi Aoki
Associate Editor: Alexandre Rasi Aoki

Received: 24-Feb-2024; Accepted: 15-Apr-2024

*Correspondence: minakshiy10@gmail.com; Tel.: +91-9025723183

HIGHLIGHTS

- Proposed a novel fused feature selection technique
- The proposed Chi-Vec captures the semantic and statistical features.
- Explored the proposed method with three distinct datasets.

Abstract: Sentiment analysis holds paramount importance in contemporary business landscapes, particularly in leveraging insights from the extensive pool of social media data. The rise of social media platforms, including opinion polls, weblogs, Twitter, and various other networks, has accentuated the need for effective sentiment analysis tools. Deep learning has emerged as a pivotal technique in natural language processing (NLP), particularly for sentiment analysis tasks, owing to its ability to autonomously learn features. However, the performance of deep learning models can suffer when confronted with a large number of features. To address this limitation, this paper proposes a novel fused feature selection technique, Chi-Vec, aimed at selectively passing relevant features to deep learning models. Chi-Vec is a fusion of Chi-square and Word2Vec. The research encompasses the exploration of three distinct datasets; CBET, ATIS, and AWARE. Leveraging the bi-directional Long Short-Term Memory (Bi-LSTM) architecture in conjunction with Chi-Vec, the approach achieves remarkable accuracy rates of 97.96%, 98.41%, and 94.45% for CBET, ATIS, and AWARE dataset respectively. Chi-Vec not only enhances the efficiency and accuracy of sentiment analysis but also demonstrates promising potential for various NLP applications.

Keywords: Sentiment Analysis; Deep Learning; Feature Selection; Chi-Vec.

INTRODUCTION

Sentiment analysis has become indispensable in modern businesses, leveraging the vast reservoir of social media data to derive actionable insights. The proliferation of online platforms has underscored the significance of understanding, and analyzing public sentiment. This burgeoning demand for sentiment analysis coincides with the exponential growth of social media, necessitating advanced techniques to sift through the immense volume of textual data and extract meaningful sentiments.

Deep learning has emerged as a powerful tool in natural language processing (NLP), revolutionizing sentiment analysis with its ability to automatically learn intricate features from raw text data. Its application in sentiment analysis has yielded remarkable results, enabling more accurate sentiment classification and sentiment-based decision-making. However, deep learning models may encounter challenges when dealing with high-dimensional feature spaces [1], prompting the need for effective feature selection techniques [2]. The deep learning models used in this research are Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), Bidirectional LSTM (Bi-LSTM), Gated Recurrent Unit (GRU), and Bidirectional Encoder Representations from Transformers (BERT).

Addressing the limitations of deep learning models, this paper introduces Chi-Vec, a novel fused feature selection technique designed to optimize sentiment analysis tasks. In this context, the integration of Chi-Square and Word2Vec features, fused together to form Chi-Vec features, holds significance. By selectively passing relevant features to deep learning architectures, Chi-Vec aims to enhance model performance and efficiency. Through the fusion of feature selection and deep learning methodologies, this approach endeavors to expand the horizons of sentiment analysis accuracy and applicability in real-world scenarios [3].

Literature Review

In their study, Sharma and Jain [4] delve into ensemble learning, evaluating multiple ensemble classifiers and features for sentiment classification of social media data. Their proposed hybrid ensemble learning model amalgamates ensemble features, namely Information Gain and CHI-Squared, with ensemble classifiers such as Ada Boost with SMO-SVM and Logistic Regression. This combination aims to enhance sentiment classification accuracy. The model exhibits notable efficacy, achieving an accuracy of 88.2% with a minimal error rate. These results signify advancements over existing state-of-the-art methods.

Alarifi and coauthors [5] proposed an innovative sentiment analysis approach leveraging big data and machine learning. Through meticulous data collection and preprocessing to minimize noise, they employed a greedy feature selection strategy to identify impactful features. These features were integrated into an optimized classifier, CSO-LSTMNN, outperforming the PSO algorithm in accuracy enhancement and error rate reduction. Evaluation metrics including error rate, precision, recall, and accuracy showcased notable enhancements in system efficiency.

Dey and Das [6] introduced a hybrid neural network-based sentiment analysis framework, enhancing TF-IDF with a non-linear global weighting factor and the k-best selection method for improved feature vectorization. Utilizing pre-trained Word2Vec embedding, the framework optimizes deep learning efficiency. Combining CNN and LSTM, it outperforms traditional machine learning methods in sentiment analysis. Validation across diverse datasets confirms its efficacy using various metrics.

Kaur and Sharma [7] developed a consumer review summarization model utilizing Natural Language Processing (NLP) and Long Short-Term Memory (LSTM) for sentiment analysis. Their hybrid approach encompasses NLP pre-processing to filter undesirable data, followed by feature extraction combining review-related and aspect-related features. Sentiment classification is then performed using LSTM. Experimental evaluation across three datasets reveals high performance, with an average precision of 94.46%, average recall of 91.63%, and average F1-score of 92.81%.

Daniel and Meena [8] emphasize the efficacy of hybrid methodologies combining deep learning and lexicon-based sentiment analysis (SA) techniques, showcasing superior performance. Their proposed approach integrates the Valence Aware Dictionary for Sentiment Reasoning (VADER) with a hybrid deep learning method, namely attention-based bidirectional long short-term memory, and variable pooling convolutional neural network (VPCNN-ABiLSTM). Achieving a remarkable accuracy of 97.1% with a 13.6% reduction in features compared to existing methods, their approach demonstrates significant advancements. The paper further discusses the CHOA approach, OBL strategy, and CNN-LSTM classifier employed within the proposed SA framework.

Overall, the existing studies demonstrate advancements in sentiment analysis methodologies, addressing challenges such as noisy data, feature selection, classifier optimization, and incorporating contextual information. However, there remain challenges in effectively handling complex data structures, improving model interpretability, and achieving consistent performance across diverse datasets. These challenges provide opportunities for further research and development in the field of sentiment analysis.

Motivation and Justification

The rapid evolution of social media platforms has generated an unprecedented volume of textual data, making sentiment analysis crucial for extracting actionable insights. While deep learning has revolutionized sentiment analysis, the challenge lies in optimizing model performance amidst high-dimensional feature spaces. Thus, the motivation for this research stems from the pressing need to enhance sentiment analysis methodologies by addressing the scalability limitations of deep learning models. By introducing Chi-Vec, a novel fused feature selection technique, the research aims to selectively pass relevant features to optimize model performance, thereby improving the efficiency and accuracy of sentiment analysis tasks across diverse datasets and applications.

This research introduces an innovative approach to sentiment analysis, addressing scalability challenges inherent in deep learning models. Chi-Vec, a fused feature selection technique, optimizes model performance and streamlines sentiment analysis processes, facilitating informed decision-making across various domains. By leveraging Chi-Square feature selection for identifying statistically significant features and Word2Vec embeddings for capturing semantic relationships, Chi-Vec enriches the feature space comprehensively. This selective extraction of relevant features aims to enhance the efficiency of deep learning architectures. Through the integration of feature selection and deep learning methodologies, Chi-Vec expands the horizons of sentiment analysis accuracy and its practical applications. Rigorous experimentation validates the robustness and effectiveness of Chi-Vec, promising to revolutionize sentiment analysis approaches and extract valuable insights from textual data in real-world scenarios.

Outline of the research work

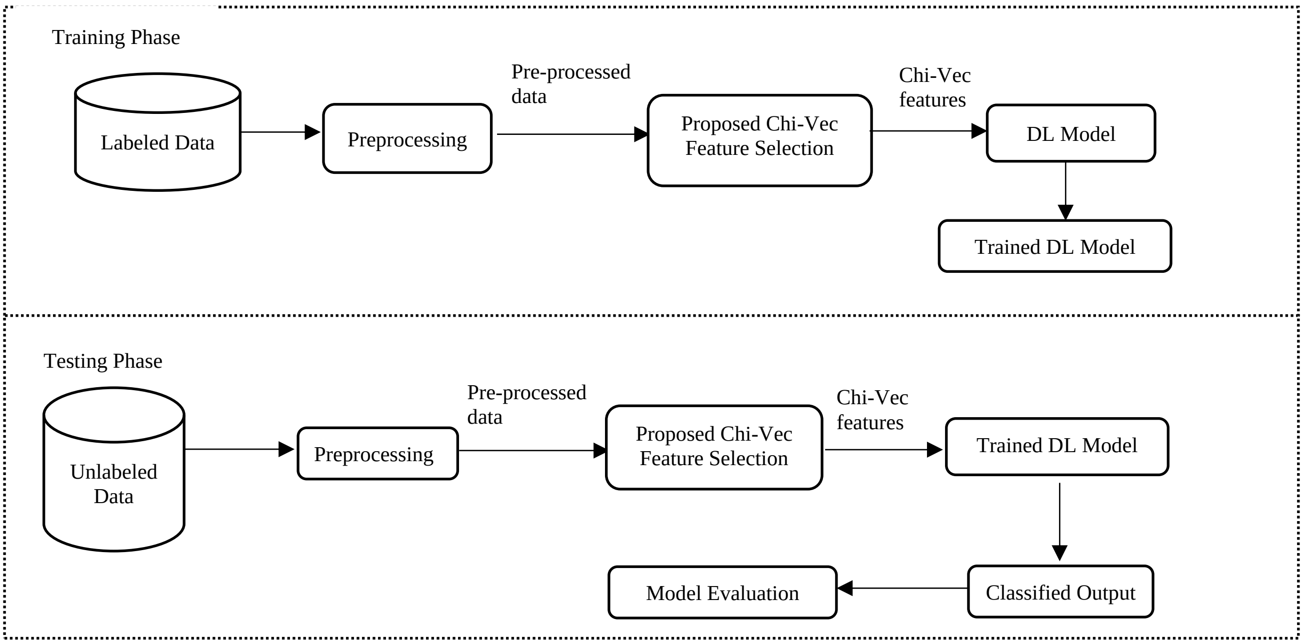

Figure 1. Outline of the research work

Figure 1 illustrates the comprehensive workflow of the research work. In the training phase, labeled data undergoes preprocessing and Chi-Vec feature selection before being fed into a Deep Learning model. Subsequently, during the testing phase, unlabeled preprocessed data is subjected to Chi-Vec feature selection and then passed through the trained Deep Learning model. Following this, the data is classified, and the model's performance is evaluated. The rest of the paper is organized as follows: Section 2 presents the methodology of the proposed work. Section 3 presents the experimental results and section 4 concludes the paper

MATERIAL AND METHODS

The proposed sentiment analysis model initiates with preprocessing steps, followed by employing a feature selection technique to identify pertinent features. Subsequently, various deep learning models are utilized to construct the sentiment analysis model.

Preprocessing

The textual data for sentiment analysis undergoes several preprocessing techniques to eliminate noise [9]. These techniques include case normalization, tokenization, lemmatization, removal of stop words, URLs, hashtags, symbols, special characters, and numbers [10,11,12,13]. Case normalization involves converting all text to either lowercase or uppercase. This helps in treating words with the same characters but different cases for e.g., "Hello" and "hello" as identical tokens. It simplifies the vocabulary and reduces sparsity in the data, making it easier for the model to learn patterns. Tokenization involves breaking down the text into individual words or tokens whereas lemmatization is the process of reducing words to their base or root form. It helps in standardizing the vocabulary by converting different inflected forms of words to their common base form. For example, "running," "runs," and "ran" would all be lemmatized to "run." Stop words are common words that often appear frequently in the text but carry little semantic meaning (e.g., "the," "is," "and"). Removing stop words reduces the dimensionality of the data and focuses the model's attention on more informative words. Additionally, to address imbalanced datasets, resampling is performed using a hybrid SMOTE-ENN technique [14].

Hybrid SMOTE-ENN

The hybrid SMOTE-ENN (Synthetic Minority Over-sampling Technique combined with Edited Nearest Neighbors) resampling technique is specifically designed to address the issue of imbalanced datasets. Imbalanced datasets occur when one class is significantly more prevalent than others, which can lead to biased models favoring the majority class. SMOTE focuses on the minority class by generating synthetic examples to balance the class distribution [15]. It does this by creating synthetic samples along the line segments connecting similar minority class instances. This helps to increase the representation of the minority class in the dataset. ENN, on the other hand, focuses on the majority class by removing instances that are considered noise or outliers. It works by identifying instances in the majority class that are misclassified by their nearest neighbors and removes them. This helps to reduce the dominance of the majority class and improve the balance between classes. By combining SMOTE and ENN, the hybrid approach aims to create a more balanced dataset that retains the essential information from both the minority and majority classes. In the case of the ATIS dataset, the class atis_flight is the majority class with 3666 instances, while the remaining five classes represent the minority classes. To address this class imbalance and ensure that the model learns from all classes equally, the hybrid SMOTE-ENN technique is applied. This approach creates a more balanced dataset, thereby enhancing the model's performance and predictive accuracy across all classes.

The rationale behind these preprocessing techniques is to create a clean and standardized dataset that retains the essential information for sentiment analysis while eliminating irrelevant noise. By applying these techniques, the model can focus on learning meaningful patterns from the text, leading to more accurate sentiment predictions.

Proposed Chi-Vec Fused Feature Selection Algorithm

The aim of this research is to propose a novel approach in feature selection for sentiment analysis, termed the Chi-Vec Fused Feature Selection Algorithm. This innovative methodology aims to combine the strengths of two distinct techniques; Chi-square and Word2Vec. Chi-square serves as a robust statistical measure, assessing the independence between features and target labels, facilitating the identification of relevant features [16]. Conversely, Word2Vec captures semantic relationships within textual data, extracting meaningful word embeddings [17,18]. By integrating these methodologies, the algorithm achieves a synergy that capitalizes on both the statistical significance conferred by Chi-square and the semantic context provided by Word2Vec embeddings. The novelty of this approach lies in its fusion of statistical significance and semantic relevance in feature selection, thereby offering a comprehensive representation of the data. Figure 2 illustrates the schematic representation of the proposed Chi-Vec Fused Feature Selection algorithm.

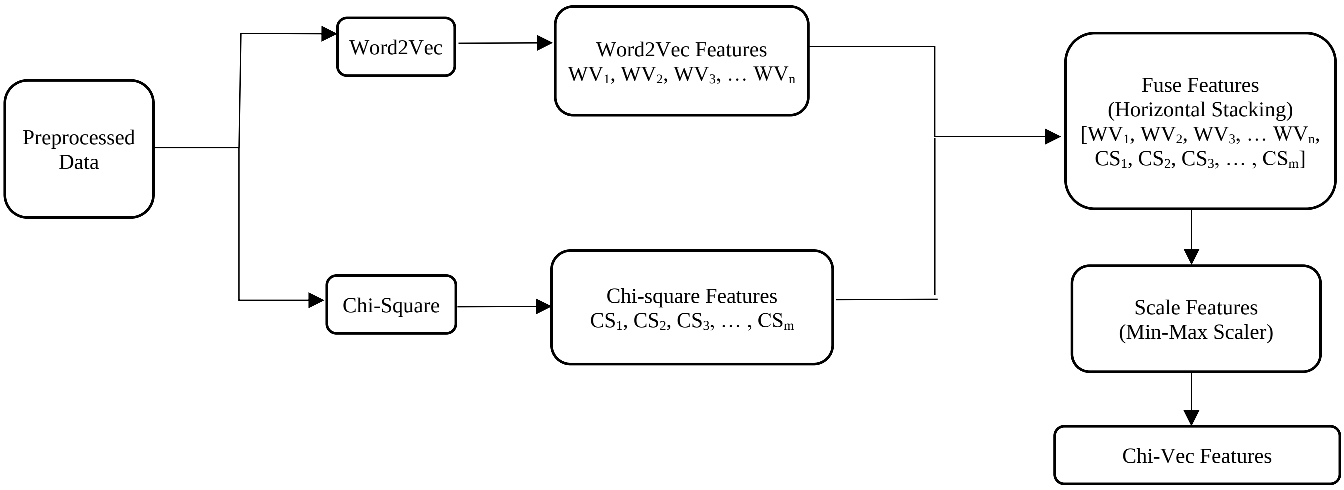

Figure 2. Working of Chi-Vec Fused Feature Selection Algorithm

Word Embeddings

To imbue the text data with semantic meaning, a pre-trained Word2Vec model google-news-300 from Gensim [19] is utilized. It produces an embedding of dimensionality 300. Initially, the text is tokenized, breaking it down into its constituent words. Subsequently, the tokens are scrutinized, retaining only those present in the Word2Vec vocabulary. For these retained tokens, word embeddings from the pre-trained Word2Vec model are accessed. These embeddings serve as numerical representations of each word's semantic context. By computing the average embedding for each text sample, a cohesive representation that encapsulates the underlying meaning and nuances of the text is created, empowering the model to comprehend the intricacies of the textual data.

Chi-Square

Chi-square feature selection is employed to identify the most influential features within the text data [1]. Initially, the textual features are encoded into numerical representations using a Keras tokenizer. These numerical features are then subjected to chi-square analysis to compute the chi-square scores. This statistical technique evaluates the relationship between each feature and the corresponding class labels, quantifying their independence with respect to class distinctions. Subsequently, chi-square scores are assigned to each feature, reflecting their significance in predicting class labels. The top k features with the highest chi-square scores are selected, facilitating the reduction of feature dimensionality while retaining essential information. This process enhances the model's ability to discern crucial patterns within the text data by focusing on the most discriminative features.

Fusion of Chi-Vec features

The fusion process begins by integrating the selected chi-square features with the Word2Vec embeddings to form a unified feature representation known as "Chi-Vec." Once both the chi-square selected features and the Word2Vec embeddings are obtained, they are combined into a single feature matrix. This fusion is accomplished through horizontal stacking, where the selected features and embeddings are concatenated horizontally. Horizontal stacking (~) involves aligning the feature vectors or matrices side by side along the horizontal axis, effectively merging them into a larger matrix. For instance, if the chi-square selected features have dimensions of [n_samples, n_features] and the Word2Vec embeddings have dimensions of [n_samples, embedding_size], horizontal stacking results in a combined feature matrix with dimensions of [n_samples, n_features ~ embedding_size].

Scaling Chi-Vec Features

To ensure uniformity and comparability across feature scales, the fused Chi-Vec features undergo feature scaling using the Min-Max Scaler [20]. This scaler transforms feature values to a predetermined range, usually between 0 and 1, while preserving the relative distribution of the data. Scaling is crucial as it prevents any single feature from disproportionately influencing the model's learning process. Without scaling, features with larger magnitude may dominate the model, leading to biased learning. By standardizing feature

scales, model is stabilized and converged, ensuring that each feature contributes proportionally to the model's decision-making process. Ultimately, scaling optimizes the model's performance during both training and inference phases.
For better understanding of this process an example is given.

Raw data:
“I like the movie. It was fantastic. It is so positive vibe.”

Preprocessed data:
“like movie fantastic positive vibe”.

Word Embedding of dimension n:
[-0.02224731 0.05386353 0.0729599 … -0.00585938 0.15139771]

Chi-square features of dimension m:
[163 522 2594 … 522 2595]

Fused Chi-Vec features of dimension (n~m):
[-0.02224731 0.05386353 0.0729599 … -0.00585938 0.15139771 163 522 2594 … 522 2595]

Scaled features:
[-0.87097477 0.43690864 0.74937503 … -1.43623798 0.70919307 0.02709956 -0.020692 … -0.01490878 -0.04417613]

Classification Techniques

This section offers a succinct overview of the diverse deep learning (DL) models utilized for classification. The models encompass CNN, RNN, LSTM, Bi-LSTM, GRU, and BERT, each contributing distinct capabilities and architectures to the classification task [21].

CNN

CNNs employ convolutional layers, pooling layers, and fully connected layers [22]. In CNN architecture, convolutional layers extract features through filters, capturing spatial hierarchies [9]. Pooling layers reduce spatial dimensions while retaining important information. Fully connected layers interpret extracted features for classification [10]. CNNs are primarily used for image analysis but are adapted for text sentiment analysis by treating text inputs as 1D sequences [20]. Word embeddings represent text, serving as input to CNNs. Through convolution and pooling, CNNs learn hierarchical text features, enabling sentiment classification. This approach efficiently captures local and global patterns, making CNNs effective for sentiment analysis in textual data.

RNN

Recurrent Neural Networks (RNNs) are a type of neural network architecture designed to process sequential data by maintaining a hidden state [9]. Each RNN unit receives an input and a hidden state from the previous time step, producing an output and updating its hidden state [17]. This recurrent structure enables RNNs to capture temporal dependencies in data, making them suitable for tasks like sentiment analysis. By processing text sequentially, RNNs can understand the context of words in a sentence, allowing them to infer sentiment [18]. This capability makes RNNs valuable in sentiment analysis applications, where understanding text sentiment is essential for decision-making.

LSTM

Long Short-Term Memory (LSTM) is a specialized recurrent neural network (RNN) architecture engineered to combat the vanishing gradient dilemma [9], offering enhanced capabilities in retaining information across lengthy sequences. Featuring memory cells equipped with input, output, and forget gates, LSTM efficiently regulates the flow of information, ensuring robust processing of sequential data [23]. Particularly in text sentiment analysis, LSTM shines by adeptly grasping the contextual intricacies within sentences, vital for discerning nuanced sentiments. Through sequential text processing, LSTM models acquire proficiency in recognizing and interpreting sentiment patterns, leveraging their innate ability to maintain pertinent information over extended sequences. This makes LSTM a preferred choice for tasks demanding a profound understanding of text sentiment across diverse contexts and lengths of textual data.

Bi-LSTM

Bidirectional Long Short-Term Memory (Bi-LSTM) is a recurrent neural network (RNN) architecture designed to capture sequential patterns bidirectionally [20]. Comprising forward and backward LSTM layers, it processes input data in both chronological orders, enhancing context comprehension [16]. In text sentiment analysis, Bi-LSTM considers the entirety of a sentence, discerning nuanced dependencies between words. By capturing long-range dependencies and contextual nuances, it excels in understanding sentiment nuances within text data [24]. Through its bidirectional nature, Bi-LSTM effectively grasps the semantic meaning of text, enabling accurate sentiment classification. Its ability to comprehend sequences bidirectionally makes it a potent tool for nuanced sentiment analysis tasks [1].

GRU

GRU, known as Gated Recurrent Unit, embodies a form of recurrent neural network architecture crafted to tackle the vanishing gradient issue inherent in traditional RNNs [9]. With reset and update gates constituting its essence, GRU selectively preserves or refreshes information, thereby enhancing learning efficacy [1]. In the realm of text sentiment analysis, GRU models meticulously process sequential input, adept at grasping subtleties inherent in contextual understanding pivotal for sentiment interpretation. Through dynamically adapting gate states, GRU adeptly captures extensive dependencies within text data, facilitating precise sentiment classification. Its streamlined framework and capacity to retain pertinent information render GRU a favored option for sentiment analysis endeavors, empowering sentiment-aware applications.

BERT

BERT utilizes a transformer architecture comprising self-attention mechanisms for contextual understanding [25,26]. It employs multiple layers of bidirectional transformers to capture intricate relationships within text [9]. BERT's pre-training involves masked language modeling and next sentence prediction tasks, enabling it to learn rich representations of words. In text sentiment analysis, BERT excels by comprehending nuances and context, facilitating more accurate sentiment classification [3]. By fine-tuning BERT on sentiment-labeled data, it adapts to specific sentiment analysis tasks, achieving superior performance compared to traditional methods. Leveraging its deep contextual understanding, BERT revolutionizes sentiment analysis, enhancing accuracy and capturing subtleties in text sentiment interpretation.

RESULTS

The study conducted a thorough experimental analysis to evaluate the efficacy of the proposed Chi-Vec feature selection algorithm against existing methods. It aimed to determine the effectiveness of various classifiers in conjunction with Chi-Vec for selecting the most relevant features in sentiment analysis tasks. Furthermore, the experiment sought to assess Chi-Vec's performance across different datasets, providing a comprehensive understanding of its applicability and potential for improving sentiment analysis accuracy.

Dataset Description

Three different benchmark datasets are used for the assessment of the Chi-Vec feature selection. They are described below.

CBET Dataset

The CBET (Cleaned Balanced Emotional Tweets) Dataset is created by Shahraki and Zaiane [27]. Its description and statistical information are listed in Table 1 and Table 2 respectively.

Table 1. Data Description of CBET Dataset

S.No.	Field	Data Type	Description
1.	Text	text	Raw text Data
2.	Emotion	text	The emotion of the text

Table 2. Statistical Information of CBET Dataset

S.No.	Class Name	No. of Instances
1.	Anger	8000
2.	Fear	8000
3.	Joy	8000
4.	Love	8000
5.	Sadness	8000
6.	Surprise	8000
7.	Thankfulness	8000
8.	Disgust	8000
9.	Guilt	8000
	Total Instances	72000

ATIS Dataset

The ATIS dataset [28] is a benchmark dataset of Airline Travel Information System created from tweets. The ATIS dataset description is listed in Table 3. Its statistical information is listed in Table 4.

Table 3. Data Description of ATIS Dataset

S.No.	Field	Data Type	Description
1.	Intent Id	text	Id of the Intent
2.	Text	text	Airline travel information text

Table 4. ATIS Dataset Statistical Information

S. No	Class Name	No. of Instances
1.	atis _ abbreviation	147
2.	atis _ aircraft	81
3.	atis _ airfare	423
4.	atis _ airline	157
5.	atis _ flight	3666
6.	atis _ flight _ time	54
7.	atis _ ground _ service	255
	Total Instances	4783

AWARE Dataset

The AWARE dataset [29] is a benchmark dataset of 11323 apps reviews. Reviews were collected from three domains: productivity, social networking, and games. The AWARE dataset description is listed in Table 5. Its statistical information is listed in Table 6. P denotes positive class and N denoted negative class.

Table 5. Data Description of AWARE Dataset

S.No.	Field	Data Type	Description
1.	review	Text	The raw text
2.	sentiment	text	The sentiment behind the review

Table 6. AWARE Dataset Statistical Information

S. No	Class Name	No. of Instances
1.	P	25000
2.	N	25000
	Total instances	50000

Performance Metrics

Table 7 presents the metrics employed to assess the performance of the classifiers. The metrics encompass Accuracy, Precision, Recall, and F1-Score [30].

Table 7. Performance Metrics and Formula

Performance Metric	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Recall	$\frac{TP}{TP+FN}$
Precision	$\frac{TP}{TP+FP}$
F1-score	$\frac{2}{\frac{1}{P}+\frac{1}{R}}$
	Where P is the precision and R is the Recall

In this context, TP denotes the count of accurately identified positive occurrences, TN denotes the count of accurately identified negative occurrences, FP signifies the count of occurrences identified as positive but are actually negative, and FN indicates the count of occurrences identified as negative but are actually positive.

Experimental Setup

The hyperparameters used in the experimental setup included input dense layer units of 64 with ReLU activation, a dropout rate of 0.5, an output dense layer and softmax activation, employing categorical crossentropy as the loss function, utilizing the Adam optimizer, running for 10 epochs with a batch size of 25, and validating with a split ratio of 0.1. These hyperparameters were carefully selected to ensure a robust evaluation framework and facilitate fair comparisons across sentiment analysis tasks and datasets.

Experiment 1: Choosing the k value for chi-square in Chi-Vec Feature selection

Table 8 help to select the best k value for chi-square in Chi-Vec feature selection to choose the relevant features.

Table 8. Choosing k value for chi-square in Chi-Vec Feature Selection

Models	k value for chi-square	Accuracy (%)		
		CBET	ATIS	AWARE
CNN	1000	85.96	86.22	83.48
	500	88.00	89.09	84.82
	200	91.47	90.72	85.73
	100	95.89	93.39	86.20
	50	93.93	96.36	85.10
	40	92.15	95.95	84.82

Cont. Table 8

RNN	1000	89.25	86.73	83.84
	500	91.26	89.91	85.59
	200	94.57	92.14	87.15
	100	96.71	94.59	88.20
	50	95.42	96.46	87.48
	40	94.75	95.68	86.96
LSTM	1000	90.96	88.76	86.86
	500	93.48	91.21	88.53
	200	96.73	93.75	91.38
	100	97.71	95.98	92.20
	50	96.54	97.36	91.39
	40	95.94	96.86	90.85
Bi-LSTM	1000	91.21	89.91	87.46
	500	93.89	90.35	91.99
	200	96.57	92.89	93.58
	100	**97.96**	96.53	**94.45**
	50	95.78	**98.41**	93.01
	40	90.43	97.99	90.97
GRU	1000	88.54	86.49	80.96
	500	91.44	88.83	83.19
	200	94.17	90.86	85.69
	100	96.31	92.47	87.00
	50	95.17	93.35	85.86
	40	94.76	92.59	84.98
BERT	1000	90.95	88.26	84.49
	500	93.00	90.92	86.67
	200	95.11	93.55	88.99
	100	96.92	95.18	90.50
	50	95.62	96.66	89.10
	40	94.99	95.97	88.69

Experiment 2: Performance evaluation of the proposed Chi-Vec Feature selection for CBET dataset

Table 9 displays DL model performance with and without feature selection, illustrating the impact of feature selection on model effectiveness for CBET dataset.

Table 9. Assessment of the effectiveness of the proposed Chi-Vec feature selection technique for CBET Data

Model	Feature Selection Techniques	Accuracy(%)	Precision(%)	Recall(%)	F-score(%)
CNN	Without Feature Selection	92.09	92.23	92.09	92.84
	Chi-square	95.14	95.58	95.14	96.19
	Mutual Information	93.84	93.98	93.84	94.59
	Word2Vec	94.84	94.98	94.84	95.59
	PCA	93.34	93.48	93.34	94.09
	Proposed Chi-Vec	95.89	96.03	95.89	96.64

Cont. Table 9

RNN	Without Feature Selection	92.91	93.00	92.91	90.73
	Chi-square	96.26	96.35	96.26	94.08
	Mutual Information	94.66	94.75	94.66	92.48
	Word2Vec	95.66	95.75	95.66	93.48
	PCA	94.16	94.25	94.16	91.98
	Proposed Chi-Vec	96.71	96.80	96.71	94.53
LSTM	Without Feature Selection	93.91	92.76	93.91	92.45
	Chi-square	96.26	96.11	96.26	95.80
	Mutual Information	95.66	94.51	95.66	94.20
	Word2Vec	96.00	95.51	96.00	95.20
	PCA	95.16	94.01	95.16	93.7
	Proposed Chi-Vec	97.71	96.56	97.71	96.25
Bi-LSTM	Without Feature Selection	94.96	93.81	94.96	93.5
	Chi-square	96.31	95.16	96.31	94.85
	Mutual Information	94.71	93.56	94.71	93.25
	Word2Vec	95.71	94.56	95.71	94.25
	PCA	94.21	93.06	94.21	92.75
	Proposed Chi-Vec	**97.96**	**97.61**	**97.96**	**97.30**
GRU	Without Feature Selection	92.51	92.37	92.51	90.14
	Chi-square	95.86	95.72	95.86	93.49
	Mutual Information	94.26	94.12	94.26	91.89
	Word2Vec	95.26	95.12	95.26	92.89
	PCA	93.76	93.62	93.76	91.39
	Proposed Chi-Vec	96.31	96.17	96.31	93.94
BERT	Without Feature Selection	93.12	92.26	93.12	89.84
	Chi-square	95.97	95.61	95.97	93.19
	Mutual Information	94.87	94.01	94.87	91.59
	Word2Vec	95.87	95.01	95.87	92.59
	PCA	94.37	93.51	94.37	91.09
	Proposed Chi-Vec	96.92	96.06	96.92	93.64

Experiment 3: Performance evaluation of the proposed Chi-Vec Feature selection for ATIS dataset

Table 10 displays DL model performance with and without feature selection, illustrating the impact of feature selection on model effectiveness for ATIS dataset.

Table 10. Assessment of the effectiveness of the proposed Chi-Vec feature selection technique for ATIS Data

Model	Feature Selection Techniques	Accuracy(%)	Precision(%)	Recall(%)	F-score(%)
CNN	Without Feature Selection	91.56	91.63	91.56	90.68
	Chi-square	94.91	94.98	94.91	94.03
	Mutual Information	93.31	93.38	93.31	92.43
	Word2Vec	94.31	94.38	94.31	93.43
	PCA	92.81	92.88	92.81	91.93
	Proposed Chi-Vec	96.36	96.43	96.36	95.48
RNN	Without Feature Selection	91.66	91.64	91.66	90.73
	Chi-square	95.01	94.99	95.01	94.08
	Mutual Information	93.41	93.39	93.41	92.48
	Word2Vec	94.41	94.39	94.41	93.48
	PCA	92.91	92.89	92.91	91.98
	Proposed Chi-Vec	96.46	96.44	96.46	95.53

Cont. Table 10

LSTM	Without Feature Selection	92.56	92.15	92.56	91.62
	Chi-square	95.91	95.5	95.91	94.97
	Mutual Information	94.31	93.9	94.31	93.37
	Word2Vec	95.31	94.9	95.31	94.37
	PCA	93.81	93.4	93.81	92.87
	Proposed Chi-Vec	97.36	96.95	97.36	96.42
Bi-LSTM	Without Feature Selection	93.61	93.2	93.61	92.67
	Chi-square	96.96	96.55	96.96	96.02
	Mutual Information	95.36	94.95	95.36	94.42
	Word2Vec	96.36	95.95	96.36	95.42
	PCA	94.86	94.45	94.86	93.92
	Proposed Chi-Vec	**98.41**	**98.00**	**98.41**	**97.47**
GRU	Without Feature Selection	88.55	87.12	88.55	85.58
	Chi-square	91.90	90.47	91.90	88.93
	Mutual Information	90.30	88.87	90.30	87.33
	Word2Vec	91.30	89.87	91.30	88.33
	PCA	89.80	88.37	89.80	86.83
	Proposed Chi-Vec	93.35	91.92	93.35	90.38
BERT	Without Feature Selection	91.86	92.47	91.86	90.92
	Chi-square	95.21	95.82	95.21	94.27
	Mutual Information	93.61	94.22	93.61	92.67
	Word2Vec	94.61	95.22	94.61	93.67
	PCA	93.11	93.72	93.11	92.17
	Proposed Chi-Vec	96.66	97.27	96.66	95.72

Experiment 4: Performance evaluation of the proposed Chi-Vec Feature selection for AWARE dataset

Table 11 displays DL model performance with and without feature selection, illustrating the impact of feature selection on model effectiveness for AWARE dataset.

Table 11. Assessment of the effectiveness of the proposed Chi-Vec feature selection technique for AWARE Data

Model	Feature Selection Techniques	Accuracy(%)	Precision(%)	Recall(%)	F-score(%)
CNN	Without Feature Selection	82.20	83.04	82.20	81.93
	Chi-square	85.55	86.39	85.55	85.28
	Mutual Information	83.95	84.79	83.95	83.68
	Word2Vec	84.95	85.79	84.95	84.68
	PCA	83.45	84.29	83.45	83.18
	Proposed Chi-Vec	86.20	87.04	86.20	85.93

Cont. Table 11

RNN	Without Feature Selection	83.90	84.87	83.90	83.63
	Chi-square	87.25	88.22	87.25	86.98
	Mutual Information	85.65	86.62	85.65	85.38
	Word2Vec	86.65	87.62	86.65	86.38
	PCA	85.15	86.12	85.15	84.88
	Proposed Chi-Vec	88.20	89.17	88.20	87.93
LSTM	Without Feature Selection	87.20	87.48	87.20	87.00
	Chi-square	90.55	90.83	90.55	90.35
	Mutual Information	88.95	89.23	88.95	88.75
	Word2Vec	89.95	90.23	89.95	89.75
	PCA	88.45	88.73	88.45	88.25
	Proposed Chi-Vec	92.20	92.48	92.20	92.00
Bi-LSTM	Without Feature Selection	88.45	88.73	88.45	88.25
	Chi-square	91.80	92.08	91.80	91.60
	Mutual Information	90.20	90.48	90.20	90.00
	Word2Vec	91.20	91.48	91.20	91.00
	PCA	89.70	89.98	89.70	89.50
	Proposed Chi-Vec	**94.45**	**94.73**	**94.45**	**94.25**
GRU	Without Feature Selection	82.20	83.54	82.20	81.77
	Chi-square	85.55	86.89	85.55	85.12
	Mutual Information	83.95	85.29	83.95	83.52
	Word2Vec	84.95	86.29	84.95	84.52
	PCA	83.45	84.79	83.45	83.02
	Proposed Chi-Vec	87.00	88.34	87.00	86.57
BERT	Without Feature Selection	85.70	86.88	85.70	85.41
	Chi-square	89.05	90.23	89.05	88.76
	Mutual Information	87.45	88.63	87.45	87.16
	Word2Vec	88.45	89.63	88.45	88.16
	PCA	86.95	88.13	86.95	86.66
	Proposed Chi-Vec	90.50	91.68	90.50	90.21

Experiment 5: Assessing the training and validation loss of the Bi-LSTM model using Chi-Vec feature selection

Across Experiment 2 to Experiment 4, it is clear that Bi-LSTM consistently excels across all three datasets when employing the Chi-Vec Fused Feature selection algorithm. The training loss and validation loss of Bi-LSTM are depicted in Figure 3(a) and Figure3(b), respectively.

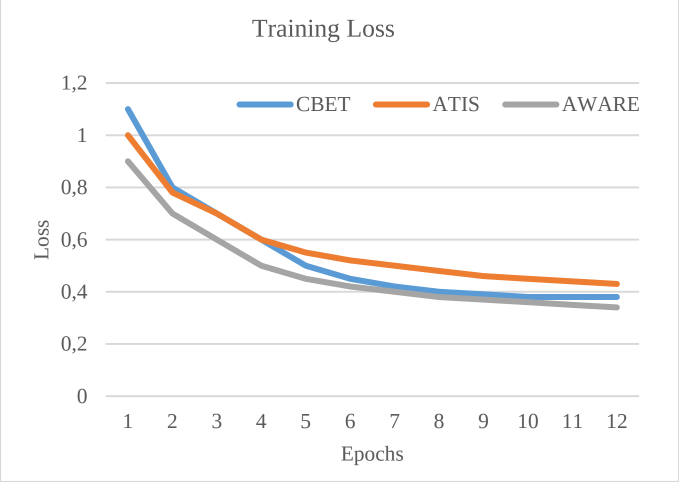

Figure 3(a). Training loss of Bi-LSTM using Chi-Vec

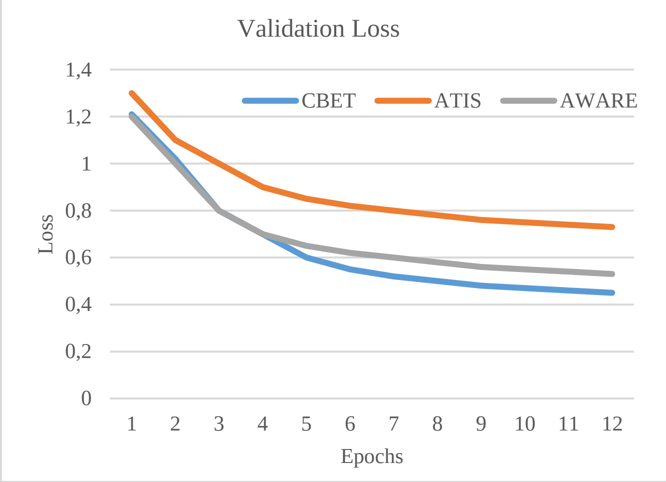

Figure 3(b). Validation Loss of Bi-LSTM using Chi-Vec

In the training and validation loss graph, the x-axis represents the number of training iterations or epochs, while the y-axis represents the loss value. Initially, both training and validation losses are typically high as the model begins learning. As training progresses, the loss generally decreases, indicating improved performance and convergence of the model. Ideally, the training loss should steadily decrease, demonstrating that the model is effectively learning from the training data. Concurrently, the validation loss should decrease as well, albeit possibly with fluctuations, indicating that the model generalizes well to unseen data. A large gap between training and validation losses suggests overfitting, while a small gap or overlap indicates good generalization thereby indicating that these graphs illustrate the model's robust performance on unseen data, indicating its generalization capabilities.

Experiment 6: Assessing Chi-Vec feature selection with state-of-art models

The proposed Chi-Vec feature selection is compared with various state-of-art models and is depicted in table 12.

Table 12. Comparison of proposed Chi-Vec feature selection with state-of-art models

Models	Accuracy (%)
SMO-SVM [4]	88.20
CSO-LSTMNN [5]	96.89
HFV+LSTM [7]	95.90
VPCNN-ABiLSTM [8]	97.10
Proposed Chi-Vec Feature Selection	**98.41**

DISCUSSION

The assessment spanning Table 9 to Table 11 demonstrates the efficacy of Chi-Vec feature selection technique across CBET, ATIS, and AWARE datasets. Chi-Vec consistently surpasses other methods, exhibiting higher accuracy, precision, recall, and F-score metrics. Remarkably, Bi-LSTM emerges as the most effective architecture across all datasets, showcasing its versatility and robustness in capturing complex patterns within the data. Bi-LSTM's ability to retain context information over long sequences helps in capturing the sentiment expressed throughout the entirety of a text. This adaptability and robustness make Bi-LSTM particularly effective in sentiment analysis tasks across diverse datasets

Figure 3(a) and Figure 3(b) validate the model's generalization to unseen data, mitigating concerns of overfitting or underfitting. Chi-Vec's success stems from its fusion of Chi-square statistics and Word2Vec embeddings, enabling it to capture both statistical and semantic features effectively. By amalgamating different feature selection techniques' strengths, Chi-Vec ensures models concentrate on pertinent information, enhancing sentiment analysis outcomes.

Moreover, this study accentuates feature selection's pivotal role in sentiment analysis, with Chi-Vec standing out as a promising approach. Its knack for selecting and representing crucial features augments model accuracy and robustness across various datasets and deep learning architectures. Overall, the findings underscore Chi-Vec's efficacy in refining sentiment analysis outcomes and propelling advancements in natural language processing. Through its adept combination of statistical and semantic features, Chi-Vec demonstrates significant potential in advancing sentiment analysis and other text-based applications, marking a noteworthy stride in the field's evolution.

CONCLUSION

This study investigated the efficacy of the Chi-Vec feature selection technique across three diverse datasets for sentiment analysis; CBET, ATIS, and AWARE. The results demonstrate a significant enhancement in sentiment analysis model performance when integrating Chi-Vec across various deep learning architectures. Consistently, across all datasets and models, Chi-Vec yielded the highest accuracy, precision, recall, and F-score compared to other feature selection techniques. Notably, the Bi-LSTM model on the ATIS dataset achieved remarkable results with Chi-Vec, boasting an accuracy of 98.41%, precision of 98.00%, recall of 98.41%, and F-score of 97.47%. Comparative analysis against Chi-square, Mutual Information, Word2Vec, and PCA further underscores the superiority of Chi-Vec in selecting discriminative features for sentiment analysis tasks. These findings underscore Chi-Vec's efficacy in improving sentiment analysis model performance, offering valuable insights for researchers and practitioners in natural language processing. Future research could investigate the scalability and generalizability of Chi-Vec across larger and more diverse datasets. Evaluating Chi-Vec's performance in multilingual sentiment analysis tasks and under different domain-specific contexts could provide deeper insights into its effectiveness across various linguistic and semantic domains.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Hussein M, Özyurt F. A new technique for sentiment analysis system based on deep learning using Chi-Square feature selection methods. Balk J Electr Comp Eng. 2021 Oct 10;9(4):320-6.
2. Kalashami MP, Pedram MM, Sadr H. EEG feature extraction and data augmentation in emotion recognition. Comput Intell Neurosci. 2022;2022(1):7028517.
3. Ma Y, He M, Pang Y, Wang L, Liu H. A multi-feature fusion approach based on domain adaptive pretraining for aspect-based sentiment analysis. Soft Comput. 2023 Nov 28:1-6.
4. Sharma S, Jain A. Hybrid ensemble learning with feature selection for sentiment classification in social media. In: Research Anthology on Applying Social Networking Strategies to Classrooms and Libraries. IGI Global; 2022. p. 1183–203.
5. Alarifi A, Tolba A, Al-Makhadmeh Z, Said W. A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. J Supercomput. 2020 Jun;76:4414-29.
6. Dey RK, Das AK. Modified term frequency-inverse document frequency based deep hybrid framework for sentiment analysis. Multimed. Tools Appl. 2023 Sep;82(21):32967-90.
7. Kaur G, Sharma A. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. J Big Data. 2023 Jan 13;10(1):5.
8. Daniel DA, Meena MJ. Deep learning-based hybrid sentiment analysis with feature selection using optimization algorithm. Multimed. Tools Appl. 2023 Apr 27:1-24.
9. Chen D, Zhengwei H, Yiting T, Jintao M, Khanal R. Emotion and sentiment analysis for intelligent customer service conversation using a multi-task ensemble framework. Cluster Comput. 2023 Jun 22:1-7.
10. Balakrishnan V, Shi Z, Law CL, Lim R, Teh LL, Fan Y. A deep learning approach in predicting products’ sentiment ratings: a comparative analysis. J. Supercomput. 2022 Apr;78(5):7206-26.
11. Dang NC, Moreno-García MN, De la Prieta F. Sentiment analysis based on deep learning: A comparative study. Electron. 2020 Mar 14;9(3):483.

12. Leelawat N, Jariyapongpaiboon S, Promjun A, Boonyarak S, Saengtabtim K, Laosunthara A, et al. Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. Heliyon. 2022 Oct 1;8(10).
13. Nandwani P, Verma R. A review on sentiment analysis and emotion detection from text. Soc Netw Anal Min 2021 Dec;11(1):81.
14. Kathuria A, Gupta A, Singla RK. AOH-Senti: Aspect-Oriented Hybrid Approach to Sentiment Analysis of Students' Feedback. SN Comput. Sci. 2023 Jan 11;4(2):152.
15. Khodaverdian Z, Sadr H, Edalatpanah SA, Nazari M. An energy aware resource allocation based on combination of CNN and GRU for virtual machine selection. Multimed. Tools Appl. 2023 Aug 23:1-28.
16. Dhal P, Azad C. A fine-tuning deep learning with multi-objective-based feature selection approach for the classification of text. Neural Comput. Appl.. 2024 Mar;36(7):3525-53.
17. Başarslan MS, Kayaalp F. Sentiment analysis using a deep ensemble learning model. Multimed. Tools Appl. 2023 Oct 16:1-25.
18. Yadav A, Vishwakarma DK. Sentiment analysis using deep learning architectures: a review. Artif. Intell. Rev.. 2020 Aug;53(6):4335-85.
19. Sadr H, Nazari M, Pedram MM, Teshnehlab M. Exploring the efficiency of topic-based models in computing semantic relatedness of geographic terms. Int. J. Web Res. 2019 Dec 1;2(2):23-35.
20. Vatambeti R, Mantena SV, Kiran KV, Manohar M, Manjunath C. Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique. Cluster Comput. 2023 Feb 1:1-7.
21. Dhanith PJ, Prabha KS. A critical empirical evaluation of deep learning models for solving aspect based sentiment analysis. Artif. Intell. Rev. 2023 Apr 6:1-60.
22. Mohades Deilami F, Sadr H, Tarkhan M. Contextualized multidimensional personality recognition using combination of deep neural network and ensemble learning. Neural Process. Lett. 2022 Oct;54(5):3811-28.
23. Thomas R, Jeba JR. A novel framework for an intelligent deep learning based product recommendation system using sentiment analysis (SA). Automatika. 2024 Apr 2;65(2):410-24.
24. Kotikam G, Selvaraj L. Golden eagle based improved Att-BiLSTM model for big data classification with hybrid feature extraction and feature selection techniques. Netw. Comput. Neural Syst. 2023 Dec 23:1-36.
25. Paramesha K, Gururaj HL, Nayyar A, Ravishankar KC. Sentiment analysis on cross-domain textual data using classical and deep learning approaches. Multimed. Tools Appl. 2023 Feb 28:1-24.
26. Sadr H, Nazari Soleimandarabi M. ACNN-TL: attention-based convolutional neural network coupling with transfer learning and contextualized word representation for enhancing the performance of sentiment classification. J. Supercomput. 2022 May;78(7):10149-75.
27. Shahraki AG, Zaiane OR. Lexical and learning-based emotion mining from text. In Proceedings of the international conference on computational linguistics and intelligent text processing 2017 (Vol. 9, pp. 24-55).
28. Hemphill CT, Godfrey JJ, Doddington GR. The ATIS spoken language systems pilot corpus. InSpeech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990
29. Alturaief N, Aljamaan H, Baslyman M. Aware: Aspect-based sentiment analysis dataset of apps reviews for requirements elicitation. In 2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW) 2021 Nov 15 (pp. 211-218). IEEE.
30. Abraham MP, KR UK. Feature based sentiment analysis of mobile product reviews using machine learning techniques. Int J Adv Trends Comput Sci Eng [Internet]. 2020;9(2):2289–96. Available from: http://dx.doi.org/10.30534/ijatcse/2020/210922020