ISSN: 0974-8571 Vol.14 No. 1 June, 2022

International Journal of Computational Intelligence in Control

Efficient Review Summarization using Sentiment Analysis: A Proposed Solution for Large Volume Reviews

Sheikh Muhammad Saqib[1], Hamid Masood Khan[1]
[1]Institute of Computing and Information Technology, Gomal University, D.I.Khan, Pakistan
Email of Author#1: saqibsheikh4@gu.edu.pk
Email of Author#2: hamidmasoodkhan@hotmail.com

Date of Submission: 30th August 2021 Revised: 12th January 2022 Accepted: 30th January 2022

Abstract: Sentiment analysis is a commonly used natural language processing technique for extracting subjective information from textual data. It has been widely used for identifying and classifying sentiment expressed in positive and negative reviews. While this technique is useful, analyzing a large volume of reviews can be time-consuming and inefficient. To address this issue, the authors propose a solution that summarizes large volumes of reviews by generating summaries from blocks of reviews using then ltk packages. The proposed methodology involves separating reviews into positive and negative categories, converting them into a number of blocks, and generating summaries for each block. The efficacy of the proposed approach is evaluated using metrics such as precision, recall, and accuracy of ROUGE scores. The results suggest that this approach is effective in helping product owners gain insights into areas for improvement and allowing customers to make informed decisions about products. This study contributes to the development of techniques that can make review analysis more efficient and effective.

Key Words: Sentiment Analysis, Machine Learning, Deep Learning, nltk-package, Rouge-Score, Text Summary

1 Introduction

Sentiment analysis is a widely used natural language processing technique that involves identifying and extracting subjective information from textual data, such as reviews, comments, and social media posts. This technique can be particularly useful in identifying and classifying the sentiment expressed in positive and negative reviews. Positive reviews typically express satisfaction and appreciation for a product or service, while negative reviews typically express dissatisfaction and criticism. Sentiment analysis algorithms can be trained on labeled data to accurately classify reviews into these two categories, allowing businesses to monitor customer feedback and identify areas for improvement. Several studies have explored the effectiveness of sentiment analysis in identifying positive and negative reviews, with some reporting high levels of accuracy [1][2].

Summary from text using machine learning is a popular natural language processing technique that involves using algorithms to automatically generate summaries of textual content. Machine learning models can be trained on large datasets of text to identify important information, such as key phrases and sentences, and generate a summary that captures the essential meaning of the original text. This technique has been applied in various domains, including news articles, legal documents, and scientific papers, to produce concise summaries that can be quickly and easily understood by users [3][4].

The majority of research in sentiment analysis has focused on analyzing reviews of products and entities. This type of analysis can be valuable not only to customers, but also to company owners. By analyzing sentiment in reviews, customers can make informed decisions about products, while companies can gain insight into the quality of their products. However, although companies can gain insight into the current perception of their product through analysis of positive and negative reviews, reading every review to identify areas for improvement may not be practical due to the large volume of reviews. To address this gap in research, the authors of this study have developed a solution to help product owners better understand issues related to their products. The generated summaries for each block exhibit a Rouge score exceeding 0.4, indicating that the accuracy of the summaries is maximized.

1.1 The study's objectives

The objective of this study is to develop a solution that summarizes large volumes of reviews to help product owners gain insights into areas for improvement. This is achieved by preparing blocks consisting of n number of reviews and generating a summary from each block.

1.2 Baseline study

The baseline of this study is to address the challenge of analyzing large volumes of reviews by developing a solution that generates summaries from blocks of reviews using the nltk package. This approach aims to help product owners gain insights into areas for improvement and better understand issues related to their products.

1.3 Problem statement

The proposed work aims to address the challenge of analyzing a large number of reviews by separating them into positive and negative categories, and then converting them into a number of blocks. The authors propose a methodology for generating summaries of positive and negative blocks using the nltk package, which can help company owners improve the quality of their products based on negative reviews and also benefit customers in making informed decisions about products. This approach eliminates the need to read every review, making the review analysis process more efficient and effective.

1.4 Research contributions

The research contributions made by this work are as follows:

Research Question 1: What methodology can be used to develop blocks from positive and negative reviews?
Research Question 2: What approach can be used to generate summaries from reviews using the nltk package?
Research Question 3: What evaluation methods can be employed to assess the performance of the proposed model and compare it with other similar studies?

2 Related Work

Filtering out objective reviews is not necessary for sentiment analysis, however doing so will also improve the precision of the analysis. There are numerous studies that examine sentence polarity in relation to figuring out the sentiment of a review or comment [5][6][7][8][9]. Sentiment orientation states that an opinion will be exactly favorable or negative depending on the situation [10]. A sentiment is a person's opinion, evaluation, or feeling about a good or service [26], feature [11][12], or both [13][14][15]. The majority of research on reviews or blogs relies on sentiment analysis with binary categorization, or positive or negative classes [16][17]. The majority of work on reviews or blogs relies on sentiment analysis with binary categorization, or positive or negative classifications [16][17]. As text categorization is carried out utilizing methodologies that are score-based, deep learning-based, and machine learning-based [18][19][20][21][22]. Machine learning and deep learning techniques employ training data, whilst other techniques use different rules based on attributes and entities. In score-based systems, orientation of opinion as favorable or unfavorable has been taken into account [20]. Work of [23][24][25] employs a combination strategy using lexical resources and SentWordNet to calculate ratings for slangs. The polarity of opinion has also been identified using a lexicon of positive and negative words using supervised [26][27][28][29] and unsupervised [30][31] approaches with increased accuracy. Latent semantic indexing has been applied to improve supervised and unsupervised methodologies in order to increase machine intelligence [32][33]. Many studies have been conducted to extract aspects and conduct aspect-based sentiment analysis in order to determine the polarity of opinions based on those aspects [34][35][36]. In addition to machine learning, deep learning has also been used extensively for sentiment analysis across a variety of dimensions [37][38][39][40][41]. In the work of [42], word2vec was utilized to reduce the number of parameters by taking a large number of words into account. Authors [43] looked into how altering convolutional neural network hyper parameters affected performance throughout numerous runs. In [44], the k-max pooling-based Op CNN model was introduced while taking the Chinese word order issue into account. The LSTM neural network was used to implement sentiment classification on tweets, identifying whether they were favorable or negative [45][46].

Prior research in this domain has employed various deep learning and machine learning techniques to classify reviews as either positive or negative. While these approaches are useful for star ranking and scoring, they do not provide a comprehensive understanding of the product's strengths and weaknesses, which are crucial for product improvement. The proposed work aims to address this research gap by filling in the missing pieces of information that have previously been overlooked.

3 Research Methodology

Figure-1 depicts the complete process for the proposed work. Initially, datasets containing reviews with positive and negative classes will be fed to a deep learning LSTM model, which will segregate positive and negative reviews based on defined review features. Subsequently, separate blocks of positive and negative reviews will be created, each block containing eight reviews. Given n-reviews in the dataset, let p and q be the numbers of positive and negative reviews, respectively. The number of blocks for positive reviews will be p/8, while the

number of blocks for negative reviews will be q/8. The lower part of Figure-1 will be utilized to generate a summary for each block.
Each block will be processed in the lower part of Figure-1 to create a frequency matrix using the frequency of unique words in each block. The maximum frequency from this block will be divided by the generated frequency matrix. The score for each review in the block will be calculated by summing the frequency of each word. The reviews in each block will be arranged based on their calculated score to generate a summary of the entire block. The statistical measure, Rouge-score, will then be calculated for each summary.
Similarly, summaries of all positive reviews blocks and negative reviews blocks will be generated to gain insight into the major issues related to the product for which the reviews were written. This process will provide a comprehensive understanding of the sentiment and issues related to the product from the available reviews.

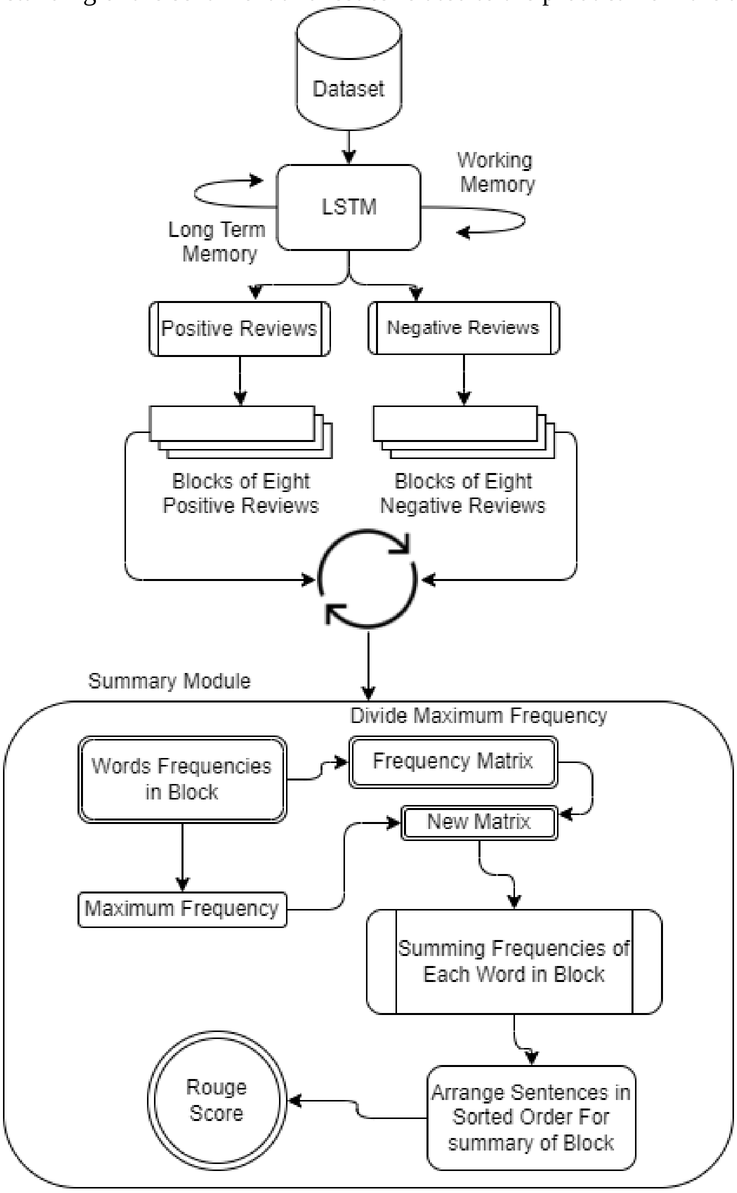

Figure 1: Proposed Work

A trained Long Short-Term Memory (LSTM) model can classify reviews into either a 'positive' or 'negative' class. The algorithm used for training the LSTM is detailed in Table-1, which employs a 100-component LSTM layer sandwiched between dense layers.

Table 1: Algorithm for LSTM Model

i)	Add Embedded Layer with 1643 neurons, as all words except stop words are 1643.
ii)	Add Dense Layer among 50 neurons with relu activation function.
iii)	Add LSTM layer with 100 units.
iv)	Add Dense Layer among 2 neurons with softmax activation function
v)	Compile Model with loss option 'categorical_ crossentropy' and optimizer as 'adam'
vi)	Fit model on 50 epochs with batch size 20.

The algorithm described above will produce two distinct lists, denoted as "positive Texts" and "negative Texts". These lists will be further divided into blocks of eight reviews each, termed "blocks Positive Strings" and "blocks Negative Strings", respectively. This division will be performed utilizing the algorithm outlined in Table-2.

Table 2: Algorithm for Generating Blocks

```
Blocks Positive Strings = []
psText = ""
for i = 1 to 160 do
   if i mod 8 == 0 then
      append ps Text to blocks Positive Strings
      set ps Textto ""
   else
      concatenate positive Texts[i] with ". " and store in ps Text
end for
blocks Negative Strings = []
ns Text = ""
for i = 161 to 320 do
   if i mod 8 == 0 then
      append nsText to blocks Negative Strings
      set nsTextto ""
   else
      concatenate negative Texts[i] with ". " and store in nsText
end for
```

The algorithm presented in Table-3 will be employed to compute the summary for each block of positive and negative reviews.

Table 3: Algorithm for Summary Generation of Each Block

```
For each Text in blocks Positive Strings:
i. Text = Preprocessing(Text)
   ii. Initialize word frequencies dictionary "word_ frequencies"
   iii. for each word in tokenized Text:
      a. If word is not in "word_ frequencies" dictionary, add it with value 1
      b. Else, increment the value of the word in "word_ frequencies" dictionary
   iv. maximum Frequency = Maximum value in "word_ frequencies" dictionary
   v. Normalize the values in "word_ frequencies" dictionary by dividing each value with maximum Frequency
   vi. Initialize dictionary "splited Sentence Scores"
   vii. for each sentence in split(Text):
      a. Initialize score for the sentence
      b. for each word Token in tokenize(sentence):
i. Add the frequency of word Token from "word_ frequencies" to score of sentence
      c. Add score of sentence to "splited Sentence Scores" for the corresponding sentence
   viii. summary Of Block = Sentence with Maximum score in "splited Sentence Scores"
   ix. Initialize Rouge scorer with options ['rouge1', 'rouge L'] and stemmer set to True
   x. R scores = scorer. score(text, summary Of Block)
   xi. Display(R scores)
```

4 Applied Example

Following the implementation of an LSTM model for segregating positive and negative reviews, a set of positively rated reviews was selected to investigate the proposed methodology. Figure-2 illustrates the procedures for generating a frequency matrix from this review block, encompassing three distinct steps as outlined in step-1, step-2, and step-3.

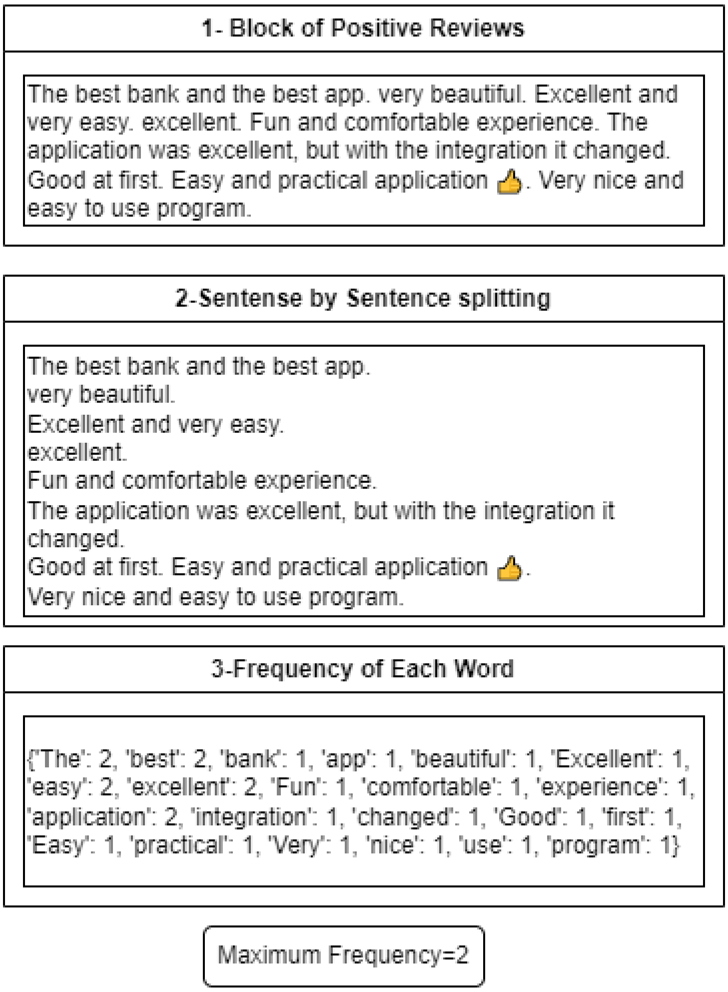

Figure 2: Frequency Matrix from Positive Block

The frequency matrix obtained in the previous step was normalized by dividing each frequency value by the maximum frequency. The resulting net matrix, with the frequencies of all words represented as a percentage, is depicted in Figure-3. Step-4 involved generating a new matrix containing the percentage frequency of each word, which was then used in step-5 to calculate the score of each sentence based on the sum of its word scores. Sentences with the highest scores were selected to form the summary. Finally, step-7 illustrates that the recall of the ROUGE score was found to be 88% and 51% for ROUGE-1 and ROUGE-L, respectively.

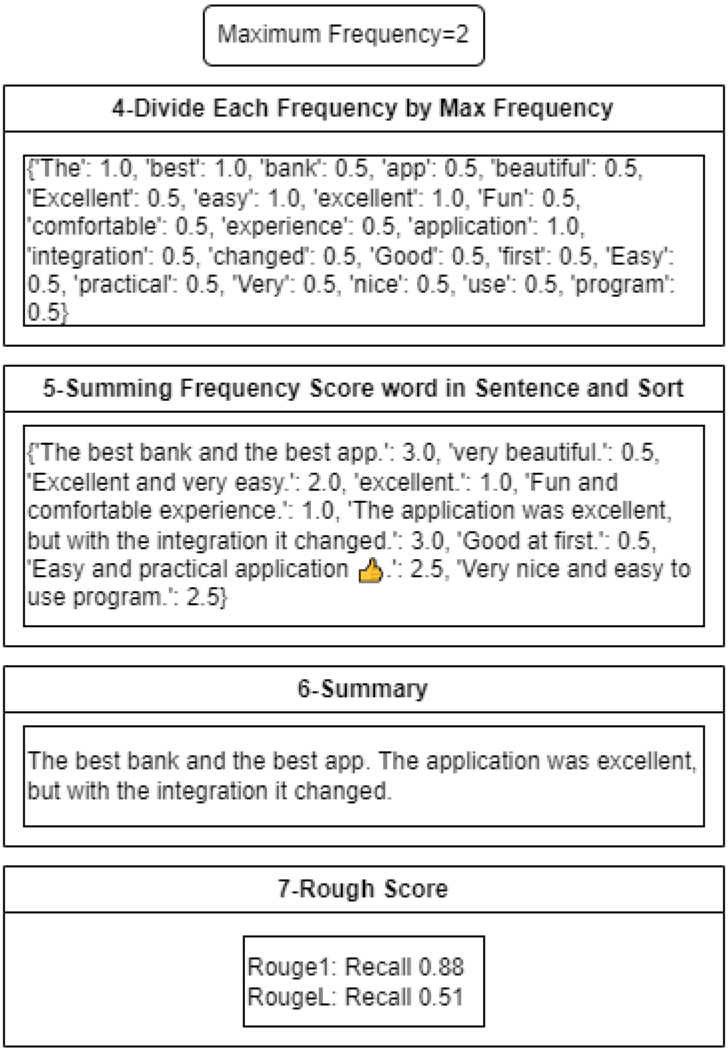

Figure 3: Summary of Given Block

5 Results and Discussion

The Kaggle platform [47]provided the Arabic_Reviews_Sentiment_analysis.csv dataset, consisting of 320 Arabic reviews, where 160 reviews were classified as positive and 160 as negative. The dataset was processed with the assistance of the google trans package [48], which was employed to translate the reviews into English. Upon classification of the reviews through the algorithm shown in Table-1, the initial 160 reviews were stored as positive and the subsequent 160 as negative. The algorithm presented in Table-2 was then utilized to generate two distinct lists named "blocks Positive Strings" and "blocks Negative Strings". The "blocks Positive Strings" list contains 20 blocks, each consisting of 160 positive reviews, whereas the "blocks Negative Strings" list contains 20 blocks of 160 negative reviews. Table-4 exhibits four examples of these blocks, where two are extracted from "blocks Positive Strings" and the other two from "blocks Negative Strings".

Table 4: Sample of Blocks

Blocks	Block’s Text
Positive Block-1	“Guaranteed services. Very nice and convenient program, thank you very much. Successful. A distinctive and concise application that facilitates the banking process. well. beautiful. Excellent time. Really very comfortable and easy to use. Above excellent.”
Positive Block-2	“The best bank and the best app. very beautiful. Excellent and very easy. excellent. Fun and comfortable experience. The application was excellent, but with the integration it changed. Good at first. Easy and practical application 👍. Very nice and easy to use program.”

Negative Block-1	"Really rubbish application. Bad app not working. Unfortunately, for the worst, the application was very wonderful, easy, and simple. Now it is complicated and does not open with you easily, and you do not find all the services available to you. The update is very slow when opening. Pass it by after the last update, it never opened, and it refuses any password you write or the card number .Al-Ahly ruined the application that most people used to rely on, unfortunately. The application does not work. Every time I log in to my account, it does not open. Look for a solution. I am disabled. Unfortunately, it is very bad. Since yesterday, I am trying to use it. I contacted customer services, and he said, delete it and download it again. It may be an update, and the same problem has not been resolved. Honestly, it's unfortunate. very bad."
Negative Block-1	"Unfortunately, it does not deserve any evaluation. God suffices them, and He is the best disposer of affairs, updating garbage. Go up, Al Ahly Bank. Technology is advancing, and you are backward. It is one of the worst programs that you can use technically. I expect its budget to be ten riyals, including tax. Unfortunately, after the update, it does not work and has many problems, and we did not benefit from it, especially outside the Kingdom. Very poor service and useless updates. Too bad the one before it is better and faster. This application does not complete the full entry and does not make transfers. Bad. puny. You have a lot of faults. The developer of the application is a failure, a failure, and the bank is a failure from it. Then you take fees and service. The old program is better, bad, and its updates do not work and are not successful. lousy. It doesn't work for me ten hours while I'm sitting on it, it doesn't work."

5.1 Summary Generation

The algorithm depicted in Table-3 was employed to produce a summary for each block by generating a frequency matrix for each one. A frequency matrix for a sample of the data is illustrated in Table-5.

Table 5: Frequency Matrix for Sample Data

Blocks	**Words' Frequencies**
Positive Block-1	{'Guaranteed': 1, 'services': 1, 'Very': 1, 'nice': 1, 'convenient': 1, 'program': 1, 'thank': 1, 'much': 1, 'Successful': 1, 'A': 1, 'distinctive': 1, 'concise': 1, 'application': 1, 'facilitates': 1, 'banking': 1, 'process': 1, 'well': 1, 'beautiful': 1, 'Excellent': 1, 'time': 1, 'Really': 1, 'comfortable': 1, 'easy': 1, 'use': 1, 'Above': 1, 'excellent': 1}
Positive Block-2	{'The': 2, 'best': 2, 'bank': 1, 'app': 1, 'beautiful': 1, 'Excellent': 1, 'easy': 2, 'excellent': 2, 'Fun': 1, 'comfortable': 1, 'experience': 1, 'application': 2, 'integration': 1, 'changed': 1, 'Good': 1, 'first': 1, 'Easy': 1, 'practical': 1, 'Very': 1, 'nice': 1, 'use': 1, 'program': 1}
Negative Block-1	{'Really': 1, 'rubbish': 1, 'application': 4, 'Bad': 1, 'app': 1, 'working': 1, 'Unfortunately': 2, 'worst': 1, 'wonderful': 1, 'easy': 1, 'simple': 1, 'Now': 1, 'complicated': 1, 'open': 2, 'easily': 1, 'find': 1, 'services': 2, 'available': 1, 'The': 2, 'update': 3, 'slow': 1, 'opening': 1, 'Pass': 1, 'last': 1, 'never': 1, 'opened': 1, 'refuses': 1, 'password': 1, 'write': 1, 'card': 1, 'number': 1, 'Al': 1, 'Ahly': 1, 'ruined': 1, 'people': 1, 'used': 1, 'rely': 1, 'unfortunately': 1, 'work': 1, 'Every': 1, 'time': 1, 'I': 4, 'log': 1, 'account': 1, 'Look': 1, 'solution': 1, 'disabled': 1, 'bad': 2, 'Since': 1, 'yesterday': 1, 'trying': 1, 'use': 1, 'contacted': 1, 'customer': 1, 'said': 1, 'delete': 1, 'download': 1, 'It': 1, 'may': 1, 'problem': 1, 'resolved': 1, 'Honestly': 1, 'unfortunate': 1}
Negative Block-2	{'Unfortunately': 2, 'deserve': 1, 'evaluation': 1, 'God': 1, 'suffices': 1, 'He': 1, 'best': 1, 'disposer': 1, 'affairs': 1, 'updating': 1, 'garbage': 1, 'Go': 1, 'Al': 1, 'Ahly': 1, 'Bank': 1, 'Technology': 1, 'advancing': 1, 'backward': 1, 'It': 2, 'one': 2, 'worst': 1, 'programs': 1, 'use': 1, 'technically': 1, 'I': 2, 'expect': 1, 'budget': 1, 'ten': 2, 'riyals': 1, 'including': 1, 'tax': 1, 'update': 1, 'work': 4, 'many': 1, 'problems': 1, 'benefit': 1, 'especially': 1, 'outside': 1, 'Kingdom': 1, 'Very': 1, 'poor': 1, 'service': 2, 'useless': 1, 'updates': 2, 'Too': 1, 'bad': 2, 'better': 2, 'faster': 1, 'This': 1, 'application': 2, 'complete': 1, 'full': 1, 'entry': 1, 'make': 1, 'transfers': 1, 'Bad': 1, 'puny': 1, 'You': 1, 'lot': 1, 'faults': 1, 'The': 2, 'developer': 1, 'failure': 3, 'bank': 1, 'Then': 1, 'take': 1, 'fees': 1, 'old': 1, 'program': 1, 'successful': 1, 'lousy': 1, 'hours': 1, 'sitting': 1}

Subsequently, each frequency value was divided by the maximum frequency in the frequency matrix to obtain a new frequency matrix. Table-6 presents the new frequency matrices for a sample of the data.

Table 6: New Matrix by the Division of Maximum Frequency

Blocks	Frequencies Divided By Max Frequency
Positive Block-1	{'Guaranteed': 1.0, 'services': 1.0, 'Very': 1.0, 'nice': 1.0, 'convenient': 1.0, 'program': 1.0, 'thank': 1.0, 'much': 1.0, 'Successful': 1.0, 'A': 1.0, 'distinctive': 1.0, 'concise': 1.0, 'application': 1.0, 'facilitates': 1.0, 'banking': 1.0, 'process': 1.0, 'well': 1.0, 'beautiful': 1.0, 'Excellent': 1.0, 'time': 1.0, 'Really': 1.0, 'comfortable': 1.0, 'easy': 1.0, 'use': 1.0, 'Above': 1.0, 'excellent': 1.0}
Positive Block-2	{'The': 1.0, 'best': 1.0, 'bank': 0.5, 'app': 0.5, 'beautiful': 0.5, 'Excellent': 0.5, 'easy': 1.0, 'excellent': 1.0, 'Fun': 0.5, 'comfortable': 0.5, 'experience': 0.5, 'application': 1.0, 'integration': 0.5, 'changed': 0.5, 'Good': 0.5, 'first': 0.5, 'Easy': 0.5, 'practical': 0.5, 'Very': 0.5, 'nice': 0.5, 'use': 0.5, 'program': 0.5}
Negative Block-1	{'Really': 0.25, 'rubbish': 0.25, 'application': 1.0, 'Bad': 0.25, 'app': 0.25, 'working': 0.25, 'Unfortunately': 0.5, 'worst': 0.25, 'wonderful': 0.25, 'easy': 0.25, 'simple': 0.25, 'Now': 0.25, 'complicated': 0.25, 'open': 0.5, 'easily': 0.25, 'find': 0.25, 'services': 0.5, 'available': 0.25, 'The': 0.5, 'update': 0.75, 'slow': 0.25, 'opening': 0.25, 'Pass': 0.25, 'last': 0.25, 'never': 0.25, 'opened': 0.25, 'refuses': 0.25, 'password': 0.25, 'write': 0.25, 'card': 0.25, 'number': 0.25, 'Al': 0.25, 'Ahly': 0.25, 'ruined': 0.25, 'people': 0.25, 'used': 0.25, 'rely': 0.25, 'unfortunately': 0.25, 'work': 0.25, 'Every': 0.25, 'time': 0.25, 'I': 1.0, 'log': 0.25, 'account': 0.25, 'Look': 0.25, 'solution': 0.25, 'disabled': 0.25, 'bad': 0.5, 'Since': 0.25, 'yesterday': 0.25, 'trying': 0.25, 'use': 0.25, 'contacted': 0.25, 'customer': 0.25, 'said': 0.25, 'delete': 0.25, 'download': 0.25, 'It': 0.25, 'may': 0.25, 'problem': 0.25, 'resolved': 0.25, 'Honestly': 0.25, 'unfortunate': 0.25}
Negative Block-2	{'Unfortunately': 0.5, 'deserve': 0.25, 'evaluation': 0.25, 'God': 0.25, 'suffices': 0.25, 'He': 0.25, 'best': 0.25, 'disposer': 0.25, 'affairs': 0.25, 'updating': 0.25, 'garbage': 0.25, 'Go': 0.25, 'Al': 0.25, 'Ahly': 0.25, 'Bank': 0.25, 'Technology': 0.25, 'advancing': 0.25, 'backward': 0.25, 'It': 0.5, 'one': 0.5, 'worst': 0.25, 'programs': 0.25, 'use': 0.25, 'technically': 0.25, 'I': 0.5, 'expect': 0.25, 'budget': 0.25, 'ten': 0.5, 'riyals': 0.25, 'including': 0.25, 'tax': 0.25, 'update': 0.25, 'work': 1.0, 'many': 0.25, 'problems': 0.25, 'benefit': 0.25, 'especially': 0.25, 'outside': 0.25, 'Kingdom': 0.25, 'Very': 0.25, 'poor': 0.25, 'service': 0.5, 'useless': 0.25, 'updates': 0.5, 'Too': 0.25, 'bad': 0.5, 'better': 0.5, 'faster': 0.25, 'This': 0.25, 'application': 0.5, 'complete': 0.25, 'full': 0.25, 'entry': 0.25, 'make': 0.25, 'transfers': 0.25, 'Bad': 0.25, 'puny': 0.25, 'You': 0.25, 'lot': 0.25, 'faults': 0.25, 'The': 0.5, 'developer': 0.25, 'failure': 0.75, 'bank': 0.25, 'Then': 0.25, 'take': 0.25, 'fees': 0.25, 'old': 0.25, 'program': 0.25, 'successful': 0.25, 'lousy': 0.25, 'hours': 0.25, 'sitting': 0.25}

The sum of each word in every sentence of a block was calculated, and then the scores were ranked to select the top-scoring sentence as the summary of the block. Table-7 illustrates the sum of scores, whereas Table-8 displays the summary.

Table 7: Summation of Scores

Blocks	**Calculated Scores of words in each sentence**
Positive Block-1	{'Guaranteed services.': 1.0, 'Very nice and convenient program, thank you very much.': 5.0, 'A distinctive and concise application that facilitates the banking process.': 6.0, 'well.': 1.0, 'beautiful.': 1.0, 'Excellent time.': 2.0, 'Really very comfortable and easy to use.': 3.0, 'Above excellent.': 1.0}
Positive Block-2	{'The best bank and the best app.': 3.0, 'very beautiful.': 0.5, 'Excellent and very easy.': 2.0, 'excellent.': 1.0, 'Fun and comfortable experience.': 1.0, 'The application was excellent, but with the integration it changed.': 3.0, 'Good at first.': 0.5, 'Easy and practical application 👍.': 2.5, 'Very nice and easy to use program.': 2.5}
Negative Block-1	{'Really rubbish application.': 1.25, 'Bad app not working.': 1.0, 'Unfortunately, for the worst, the application was very wonderful, easy, and simple.': 2.25, 'Now it is complicated and does not open with you easily, and you do not find all the services available to you.': 2.0, 'The update is very slow when opening.': 1.25, 'The application does not work.': 1.25, 'Every time I log in to my account, it does not open.': 1.25, 'Look for a solution.': 0.25, 'I am disabled.': 0.25, 'Unfortunately, it is very bad.': 0.75, 'Since yesterday, I am trying to use it.': 0.75, 'I contacted customer services, and he said, delete it and download it again.': 1.75,

	'It may be an update, and the same problem has not been resolved..': 1.5, "Honestly, it's unfortunate.": 0.25, 'very bad.': 0.5}
Negative Block-2	{'Unfortunately, it does not deserve any evaluation.': 0.5, 'God suffices them, and He is the best disposer of affairs, updating garbage.': 1.5, 'Go up, Al Ahly Bank.': 0.25, 'Technology is advancing, and you are backward.': 0.5, 'It is one of the worst programs that you can use technically.': 1.5, 'I expect its budget to be ten riyals, including tax.': 1.75, 'Unfortunately, after the update, it does not work and has many problems, and we did not benefit from it, especially outside the Kingdom.': 2.5, 'Very poor service and useless updates.': 1.5, 'Too bad the one before it is better and faster.': 1.75, 'This application does not complete the full entry and does not make transfers.': 1.75, 'Bad.': 0.5, 'puny.': 0.25, 'You have a lot of faults.': 0.5, 'The developer of the application is a failure, a failure, and the bank is a failure from it.': 3.25, 'Then you take fees and service.': 1.0, 'The old program is better, bad, and its updates do not work and are not successful.': 3.25, 'lousy.': 0.25, "It doesn't work for me ten hours while I'm sitting on it, it doesn't work.": 3.0}

Table 8: Summary of each Block

Blocks	Summary
Positive Block-1	“A distinctive and concise application that facilitates the banking process. Very nice and convenient program, thank you very much.”
Positive Block-2	“The best bank and the best app. The application was excellent, but with the integration it changed. Easy and practical application 👍.”
Negative Block-1	“Unfortunately, for the worst, the application was very wonderful, easy, and simple. Now it is complicated and does not open with you easily, and you do not find all the services available to you.”
Negative Block-2	“The developer of the application is a failure, a failure, and the bank is a failure from it. The old program is better, bad, and its updates do not work and are not successful.”

These summaries of 8 reviews in a block are advantageous concerning product improvement, rather than perusing eight individual reviews separately.

5.2 Rouge Score

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a set of metrics used for evaluating the quality of a text summary compared to a reference summary or set of summaries. ROUGE calculates various measures of overlap, such as precision, recall, and F1 score, between the n-gram units (words, phrases, or sentences) in the generated summary and those in the reference summaries. These measures are designed to capture the relevance and completeness of the generated summary. ROUGE scores are commonly used in summarization research to compare the performance of different summarization models or to tune the hyper parameters of a summarization algorithm. A ROUGE score greater than 40 is typically regarded as the best score, and in this case, the proposed model generated summaries achieved a score greater than 40, suggesting that the generated summaries are of excellent quality. The scores for the sample data can be found in Table-9.

Table 9: ROUGE-Score of Sample Data

Blocks	Rouge1: Precision	Rouge1: Recall	Rouge1: F1-Score	Rouge L: Precision	Rouge L: Recall	Rouge L: F1-Score
Positive Block-1	1.0	0.914285714	0.955223880	0.5625	0.5142857142	0.53731343283
Positive Block-2	1.0	0.880952380	0.936708860	0.7567567567	0.6666666666	0.70886075949
Negative Block-1	1.0	0.496688741	0.663716814	0.8133333333	0.4039735099	0.53982300884
Negative Block-2	1.0	0.624277456	0.768683274	0.5092592592	0.3179190751	0.39145907473

5.3 Comparison with Benchmarks

The majority of studies in this field have relied on deep learning methods for sentiment analysis, which involve classifying reviews into positive or negative categories. The improvement in the ROUGE-1 scores from 0.3788 to 0.5190 [49]is statistically significant, summaries generated from proposed work have Rouge-1 score between 0.50 to 0.90. The latest benchmark studies on sentiment analysis solely based on these techniques are presented below:

5.3.1. Comparison with Study-1 [2]

Study-1 focuses on the use of convolutional neural networks (CNNs) for sentiment analysis, which have shown promising results in natural language processing. The study analyzes the impact of various hyper-parameters, such as word embeddings, activation functions, filter sizes, feature maps, pooling methods, regularization constants, and dropout rates on the performance of the CNN model. The main motivation for this study is the need for efficient sentiment analysis classifiers to analyze large volumes of opinions and reviews on social media. The results of the study show that the proposed CNN model with optimized hyper-parameters outperforms traditional machine learning models and achieves similar performance to state-of-the-art models with gains of up to 2% in some cases.

Proposed work provides a solution to summarize large volumes of reviews using the nltk packages by generating summaries from blocks of reviews. The methodology involves separating reviews into positive and negative categories, converting them into a number of blocks, and generating summaries for each block. The study evaluates the efficacy of this approach using metrics such as precision, recall, and accuracy of ROUGE scores. The study highlights the need for efficient and effective review analysis techniques that can help product owners gain insights into areas for improvement and enable customers to make informed decisions about products.

Based on the comparison of the two studies, Study-2 is better suited for practical application as it offers an efficient solution to summarize large volumes of reviews, which can be particularly useful for businesses with a high volume of customer feedback. Additionally, the proposed approach in Study-2 is more straightforward and requires fewer hyper-parameters than the CNN model in Study-1, making it more accessible to a wider audience.

5.3.2. Comparison with Study-2[50]

Both Study-2 and propose work discuss the application of sentiment analysis in extracting subjective information from textual data. Study-2 specifically focuses on sentiment analysis of Twitter data and provides a comprehensive analysis of the most standard and widely applicable opinion mining techniques based on machine learning and lexicon-based approaches. Meanwhile, proposed work proposes a solution for summarizing large volumes of reviews by generating summaries from blocks of reviews using the nltk packages, which can be useful for product owners to gain insights into areas for improvement and for customers to make informed decisions about products.

The proposed solution is more practical and useful for real-world applications because it addresses the issue of analyzing a large volume of reviews, which can be time-consuming and inefficient. By summarizing reviews into blocks, the proposed approach can provide an overview of the sentiment expressed in the reviews and help identify areas for improvement. Additionally, the evaluation metrics used in this work, such as precision, recall, and accuracy of ROUGE scores, provide a quantitative measure of the effectiveness of the proposed approach, which can be useful for benchmarking against other techniques. Overall, proposed work provides a more practical solution for sentiment analysis of reviews and can have significant applications in product improvement and customer satisfaction.

6 Conclusion

Sentiment analysis is a widely used natural language processing technique that involves identifying and extracting subjective information from textual data. It has been particularly useful in identifying and classifying the sentiment expressed in positive and negative reviews. Sentiment analysis algorithms can be trained on labeled data to accurately classify reviews into these two categories, allowing businesses to monitor customer feedback and identify areas for improvement. Machine learning models can be trained on large datasets of text to identify important information and generate a summary that captures the essential meaning of the original text. While the majority of research in sentiment analysis has focused on analyzing reviews of products and entities, the authors of the proposed work have developed a solution to help product owners better understand issues related to their products by summarizing large volumes of reviews. The methodology involves separating reviews into positive and negative categories, converting them into a number of blocks, and generating summaries for each block. The proposed approach can provide an overview of the sentiment expressed in the reviews and help identify areas for improvement. The evaluation metrics used in this work provide a quantitative measure of the effectiveness of the proposed approach, which can be useful for benchmarking against other techniques. The proposed solution is more practical and useful for real-world applications because it addresses the issue of analyzing a large volume of reviews, which can be time-consuming and inefficient.

Additionally, the evaluation metrics used in this work provide a quantitative measure of the effectiveness of the proposed approach, which can be useful for benchmarking against other techniques. The proposed approach is more straightforward and requires fewer hyper-parameters than other techniques, making it more accessible to a wider audience.

In conclusion, sentiment analysis is a valuable technique that can be used to extract subjective information from textual data, such as reviews, comments, and social media posts. The proposed work provides a practical solution for sentiment analysis of reviews and can have significant applications in product improvement and customer satisfaction. By summarizing reviews into blocks, the proposed approach can provide an overview of the sentiment expressed in the reviews and help identify areas for improvement. The evaluation metrics used in this work provide a quantitative measure of the effectiveness of the proposed approach, which can be useful for benchmarking against other techniques.

In natural language processing, ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is a set of metrics used to evaluate the quality of summaries produced by machine learning models. ROUGE-1 is a metric that measures the overlap between unigrams (single words) in the generated summary and those in the reference summary (the original text).The statement "The improvement in the ROUGE-1 scores from 0.3788 to 0.5190 is statistically significant[49]" indicates that there was a significant improvement in the accuracy of the generated summaries. This improvement can be attributed to the proposed approach used in the study, which involved generating summaries from blocks of reviews using the nltk packages. The increase in the ROUGE-1 score from 0.3788 to 0.5190 indicates that there was a significant increase in the overlap between the unigrams in the generated summary and those in the reference summary, indicating that the generated summary accurately captures the essence of the original text.

7 Limitations and Future Work

While the proposed work has achieved promising results in summarizing product/location reviews written in English, it is important to acknowledge the various challenges of natural language processing that can lead to ambiguity in understanding sentences, even for humans. In order to improve and build upon the proposed work, it is important to consider the following limitations:

-The research is limited to product/location reviews written in English, which may not generalize well to other languages or domains. Future investigations can explore extending the proposed approach to other languages and domains.

-The proposed model divides the reviews into blocks of eight reviews, which may not be the optimal size for all types of reviews. Further research can investigate the impact of varying the block size on the accuracy of the generated summaries.

-The proposed work uses the nltk package with a frequency matrix of words to generate summaries. While this approach has shown to be effective, other algorithms can also be explored for summary generation, such as deep learning models or graph-based approaches.

Overall, while the proposed work has made a valuable contribution in summarizing reviews, there is still room for improvement and future investigations can further explore and address the limitations of the proposed approach.

8 References

[1] M. Iram, S. U. Rehman, S. Shahid, and S. A. Mehmood, “Anatomy of Sentiment Analysis of Tweets Using Machine Learning Approach,” in *Proceedings of the Pakistan Academy of Sciences*, 2022, vol. 59, no. January, pp. 63–75.

[2] N. Nedjah, I. Santos, and L. de Macedo Mourelle, “Sentiment analysis using convolutional neural network via word embeddings,” *Evol. Intell.*, vol. 15, no. 4, pp. 2295–2319, 2022, doi: 10.1007/s12065-019-00227-4.

[3] S. Urologin, “Sentiment analysis, visualization and classification of summarized news articles: A novel approach,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 616–625, 2018, doi: 10.14569/ijacsa.2018.090878.

[4] L. M. Al Qassem, D. Wang, Z. Al Mahmoud, H. Barada, A. Al-Rubaie, and N. I. Almoosa, “Automatic Arabic Summarization: A survey of methodologies and systems,” *Procedia Comput. Sci.*, vol. 117, pp. 10–18, 2017, doi: 10.1016/j.procs.2017.10.088.

[5] T. Chen, R. Xu, Y. He, and X. Wang, “Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,” *Expert Syst. Appl.*, vol. 72, pp. 221–230, 2017, doi: 10.1016/j.eswa.2016.10.065.

[6] F. Wu, Y. Huang, and Z. Yuan, “Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources,” *Inf. Fusion*, vol. 35, pp. 26–37, 2017, doi: 10.1016/j.inffus.2016.09.001.

[7] T. Naeem, F. M. Kundi, and S. M. Saqib, “Extracting Words’ Polarity With Definition and Examples,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 15, no. 2, pp. 180–190, 2017, doi: 10.1036/0071395415.

[8] wang maoquan, C. Shiyun, X. Yufei, and Z. Lu, "EICA at SemEval-2017 Task 4: A Simple Convolutional Neural Network for Topic-based Sentiment Classification," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 2017, pp. 737–740, [Online]. Available: http://www.aclweb.org/anthology/S17-2124.

[9] D. Bollegala, T. Mu, and J. Y. Goulermas, "Cross-Domain Sentiment Classification Using Sentiment Sensitive Embeddings," in *IEEE Transactions on Knowledge and Data Engineering*, 2016, vol. 28, no. 2, pp. 398–410, doi: 10.1109/TKDE.2015.2475761.

[10] B. Liu, *Sentiment Analysis and Opinion Mining*, vol. 5, no. 1. 2012.

[11] J. Jin, P. Ji, and R. Gu, "Identifying comparative customer requirements from product online reviews for competitor analysis," *Eng. Appl. Artif. Intell.*, vol. 49, pp. 61–73, 2016, doi: 10.1016/j.engappai.2015.12.005.

[12] L. Shu, H. Xu, and B. Liu, "Lifelong Learning CRF for Supervised Aspect Extraction," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017, pp. 148–154, doi: 10.18653/v1/P17-2023.

[13] K. Khan, B. B. Baharudin, A. Khan, and Fazal-E-Malik, "Mining opinion from text documents: A survey," *2009 3rd IEEE Int. Conf. Digit. Ecosyst. Technol. DEST '09*, vol. 3, no. 7, pp. 217–222, 2009, doi: 10.1109/DEST.2009.5276756.

[14] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/1500000011.

[15] M. Z. Asghar, A. Khan, S. Ahmad, and F. M. Kundi, "A Review of Feature Extraction in Sentiment Analysis," *J. Basic. Appl. Sci. Res*, vol. 4, no. 3, pp. 181–186, 2014.

[16] S. Wang, D. Li, Y. Wei, and H. Li, "A Feature Selection Method Based on Fisher 's Discriminant Ratio for Text Sentiment Classification Suge," *Springer-Verlag Berlin Heidelb.*, pp. 88–97, 2009.

[17] A. A. A. Esmin, R. L. De Oliveira, and S. Matwin, "Hierarchical classification approach to emotion recognition in twitter," in *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012*, 2012, vol. 2, no. March, pp. 381–385, doi: 10.1109/ICMLA.2012.195.

[18] W. Y. and L. H. A. Wang S, Li D, Song X, "A feature selection method based on improved fisher's discriminant ratio for text sentiment classification," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8696-8702., 2011.

[19] L. C. H. and C. H. Chen LS, "A neural network based approach for sentiment classification in the blogosphere.," *J. Informetr.*, vol. 5, no. 2, pp. 313–322, 2011.

[20] F. M. Kundi, A. Khan, S. Ahmad, and M. Z. Asghar, "Lexicon-Based Sentiment Analysis in the Social Web," *J. Basic. Appl. Sci. Res*, vol. 4, no. 6, pp. 238–248, 2014.

[21] J. Fiaidhi, O. Mohammed, S. Mohammed, S. Fong, and T. Kim, "Opinion Mining over Twitterspace : Classifying Tweets Programmatically using the R Approach," *Digit. Inf. Manag. (ICDIM), Seventh Int. Conf. on. IEEE*, pp. 313–319, 2012.

[22] A. Jeyapriya and C. S. K. Selvi, "Extracting aspects and mining opinions in product reviews using supervised learning algorithm," in *2nd International Conference on Electronics and Communication Systems, ICECS 2015*, 2015, pp. 548–552, doi: 10.1109/ECS.2015.7124967.

[23] M. Z. A. Fazal Masud Kundi, Shakeel Ahmad, Aurangzeb Khan, "Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet Fazal," *Life Sci. J.*, vol. 11, no. 1, pp. 66–72, 2014.

[24] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," *8th Int. Work. Semant. Eval. SemEval 2014 - co-located with 25th Int. Conf. Comput. Linguist. COLING 2014, Proc.*, vol. 4, no. 8, pp. 27–35, Aug. 2014, doi: 10.3115/v1/s14-2004.

[25] Deepak Kumar Gupta and Asif Ekbal, "Supervised Machine Learning for Aspect based Sentiment Analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, pp. 319–323.

[26] T. Shaikh and D. Deshpande, "A Review on Opinion Mining and Sentiment Analysis," in *IJCA Proceedings on National Conference on Recent Trends in Computer Science and Information Technology*, 2016, no. 2, pp. 6–9.

[27] A. Alghunaim, M. Mohtarami, S. Cyphers, and J. Glass, "A Vector Space Approach for Aspect Based Sentiment Analysis," *Proc. NAACL-HLT 2015*, pp. 116–122, 2015.

[28] M. Cuadros, S. Sebastian, G. Rigau, E. H. Unibertsitatea, and S. Sebastian, "V3: Unsupervised Aspect Based Sentiment Analysis for SemEval-2015 Task 12," no. SemEval, pp. 714–718, 2015.

[29] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4 : Sentiment Analysis in Twitter," *Proc. 11th Int. Work. Semant. Eval.*, vol. 3, no. 4, pp. 502–518, 2017.

[30] S. M. Street and B. Liu, "Identifying Noun Product Features that Imply Opinions," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics Human Language Technologies*, 2011, no. Issue: 2009, pp. 575–580, [Online]. Available: http://www.aclweb.org/anthology-new/P/P11/P11-2101.pdf.

[31] S. S. Htay and K. T. Lynn, "Extracting product features and opinion words using pattern knowledge in customer reviews," *Sci. World J.*, 2013, doi: 10.1155/2013/394758.

[32] H. J. Alyamani, S. Ahmad, A. H. Syed, S. M. Saqib, and Y. D. Al-Otaibi, "LSI Authentication-Based Arabic to English Text

Converter," *Int. J. Fuzzy Log. Intell. Syst.*, vol. 21, no. 4, pp. 409–422, 2021, doi: 10.5391/IJFIS.2021.21.4.409.

[33] A. O. Shakeel Ahmad, Sheikh Muhammad Saqib, "LSI Based Search Technique: Using Extracted Keywords and Key Sentences," *VAWKUM Trans. Comput. Sci.*, vol. 14, no. 2, pp. 1–8, 2017.

[34] S. M. Saqib and F. M. Kundi, "Semi Supervised Method for Detection of Ambiguous Word and Creation of Sense : Using WordNet," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 11, pp. 353–359, 2018.

[35] L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie review mining and summarization," in *Proceedings of the 15th ACM international conference on Information and knowledge management - CIKM '06*, 2006, pp. 43–50, doi: 10.1145/1183614.1183625.

[36] S. Gojali and M. L. Khodra, "Aspect based sentiment analysis for review rating prediction," 2016, doi: 10.1109/ICAICTA.2016.7803110.

[37] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electron.*, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030483.

[38] P. Cen, K. Zhang, and D. Zheng, "Sentiment Analysis Using Deep Learning Approach," *J. Artif. Intell.*, vol. 2, no. 1, pp. 17–27, 2020, doi: 10.32604/jai.2020.010132.

[39] S. Muhammad Saqib, S. Ahmad, A. Hassan Syed, T. Naeem, and F. Mazaed Alotaibi, "Grouping of Aspects into Relevant Category Based on WordNet Definitions," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 2, p. 113, 2019.

[40] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, and A. Hussain, "Sentiment analysis of persian movie reviews using deep learning," *Entropy*, vol. 23, no. 5, pp. 1–16, 2021, doi: 10.3390/e23050596.

[41] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, 2018, doi: 10.1002/widm.1253.

[42] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," in *NAACL HLT 2015 - 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, 2015, no. 2011, pp. 103–112, doi: 10.3115/v1/n15-1011.

[43] Y. Zhang and B. Wallace, "A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification," in *Pro_ceedings of the 8th International Joint Conference on Natural Language Processing*, 2015, pp. 253–263, [Online]. Available: http://arxiv.org/abs/1510.03820.

[44] S. Zhao, Z. Xu, L. Liu, M. Guo, and J. Yun, "Towards Accurate Deceptive Opinions Detection Based on Word Order-Preserving CNN," *Math. Probl. Eng.*, vol. 2018, pp. 1–8, 2018, doi: 10.1155/2018/2410206.

[45] X. Wang, Y. Liu, C. Sun, B. Wang, and X. Wang, "Predicting polarities of tweets by composing word embeddings with long short-Term memory," in *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference*, 2015, vol. 1, pp. 1343–1353, doi: 10.3115/v1/p15-1130.

[46] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, 2015, no. September, pp. 1422–1432, doi: 10.18653/v1/d15-1167.

[47] "https://www.kaggle.com/datasets/abanoubsamir004/arabic-reviews-sentiment-analysis?resource=download." .

[48] "https://github.com/ssut/py-googletrans." .

[49] R. Paulus, C. Xiong, and R. Socher, "A deep reinforced model for abstractive summarization," in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018, no. i, pp. 1–12.

[50] Gambhir M. and Gupta V., "Recent automatic text summarization techniques: A survey," *Artif. Intell. Rev.*, vol. 47, pp. 1–66, 2016.