plagarism_check.docx

by Vijay R R

Submission date: 03-May-2024 04:36PM (UTC+0530)
Submission ID: 2369766930
File name: plagarism_check.docx (646.2K)
Word count: 2878
Character count: 18780

E-commerce Product Sentiment Assessment and Aspect Analysis

Abstract:
This research paper presents a comprehensive methodology for aspect-based sentiment analysis and text classification using advanced natural language processing techniques and deep learning models. The methodology encompasses preprocessing steps such as text normalization, tokenization, lemmatization, and stop-word removal, followed by the creation of a bag-of-words representation and calculation of TF-IDF values. Additionally, rule-based aspect extraction methods and LDA-based topic modeling are employed to identify and extract aspects from text data. The integration of FastText models for similarity calculations enhances aspect identification and sentiment analysis. A custom LSTM-based model is trained for text classification, with K-fold cross-validation ensuring robust evaluation. The paper details the training process, model architecture, and performance visualization using Matplotlib. In the inference phase, the model predicts sentiment associated with identified aspects, providing insights into the sentiment analysis process. This methodology offers a structured framework for extracting aspects, predicting sentiments, and enhancing text classification accuracy in sentiment analysis tasks.

Keywords: Aspect-based Sentiment Analysis, Text Classification, Natural Language Processing, Deep Learning Models, Preprocessing Techniques, Rule-based Aspect Extraction, LDA-based Topic Modeling, FastText Models, LSTM-based Model, Sentiment Analysis.

1 Introduction

In the era of vast digital content and user-generated opinions, sentiment analysis plays a pivotal role in understanding the sentiments and attitudes expressed in textual data. Aspect-based sentiment analysis, a specialized form of sentiment analysis, delves deeper into the specific aspects or features of products, services, or entities that influence sentiment expressions. By identifying and analyzing these aspects, businesses and researchers can gain valuable insights into customer opinions, product strengths, and areas for improvement.

Traditional sentiment analysis approaches often focus on overall sentiment polarity, categorizing text as positive, negative, or neutral. However, such approaches overlook the nuanced aspects within the text that contribute to the overall sentiment. Aspect-based sentiment analysis addresses this limitation by considering the sentiment of individual aspects or features mentioned in the text.

This research paper aims to explore and implement a methodology for aspect-based sentiment analysis that integrates advanced natural language processing techniques, rule-based aspect extraction methods, topic modeling, and deep learning models. By combining these methodologies, the research endeavors to enhance the accuracy and granularity of sentiment analysis, providing a more detailed understanding of sentiment expressions in textual data.

The methodology outlined in this paper encompasses a multi-step process starting from text preprocessing, including normalization, tokenization, and lemmatization, to more advanced techniques such as rule-based aspect extraction and LDA-based topic modeling. These techniques facilitate the identification and extraction of key aspects from text data, enabling a more comprehensive analysis of sentiment.

Furthermore, the incorporation of FastText models for similarity calculations enhances the aspect identification process by capturing semantic relationships between words and aspects. The utilization of a custom LSTM-based model for text classification enables the prediction of sentiment associated with identified aspects, offering a deeper insight into sentiment analysis tasks.

Through K-fold cross-validation and performance visualization, the research ensures the robust evaluation of the trained models, validating their effectiveness in sentiment analysis tasks. The combination of these methodologies and techniques provides a structured framework for aspect-based sentiment analysis, contributing to the advancement of sentiment analysis methodologies and applications.

By delving into the intricate details of textual data, extracting relevant aspects, and predicting sentiment associated with those aspects, this research aims to provide valuable insights for businesses, researchers, and practitioners seeking to understand and leverage sentiment analysis in their domains. The following sections will delve into the methodology, results, and implications of the proposed approach, offering a comprehensive analysis of aspect-based sentiment analysis in textual data.

2 Problem Statement

Despite the advancements in sentiment analysis techniques, traditional approaches often fall short in capturing the nuanced aspects of sentiment expressed in textual data. Current sentiment analysis methods typically focus on overall sentiment polarity without considering the specific aspects or features that influence sentiment.

This limitation poses a challenge for businesses and researchers seeking to extract detailed insights from textual data, especially in domains where understanding specific aspects of sentiment is crucial, such as product reviews, customer feedback, and social media posts. The lack of granularity in sentiment analysis hinders the ability to identify key factors driving positive or negative sentiments and limits the actionable insights that can be derived from sentiment data.

The problem statement addressed in this research project is the need for a more sophisticated and granular approach to sentiment analysis: aspect-based sentiment analysis. This approach aims to overcome the shortcomings of traditional sentiment analysis by identifying and analyzing specific aspects or features within text that contribute to sentiment expressions.

By focusing on aspect-level sentiment analysis, the research seeks to enhance the accuracy, depth, and interpretability of sentiment analysis results. The challenge lies in developing a methodology that can effectively extract and analyze aspects from text data, predict sentiment associated with these aspects, and provide actionable insights for decision-making in various domains.

The project aims to tackle this problem by integrating advanced natural language processing techniques, rule-based aspect extraction methods, topic modeling, and deep learning models to create a robust framework for aspect-based sentiment analysis. Through this approach, the research endeavors to provide a comprehensive solution to the limitations of traditional sentiment analysis methods and empower businesses and researchers with a more insightful and detailed understanding of sentiment expressions in textual data.

.

3 Related Work

Lasod & Pawar (in [1]) applied Support Vector Machines (SVM) to classify sentiment in smartphone product reviews. They achieved high accuracy (94.63%) using a combination of text pre-processing, TF-IDF weighting, and SVM classification. This work highlights the effectiveness of SVM for sentiment analysis tasks.

Huaqian He's team (in [2]) proposed a fusion sentiment analysis method to explore E-commerce product experience based on customer reviews. Their approach utilizes a multi-step process involving text pre-processing, sentiment feature extraction, text vectorization, and dimensionality reduction. This method combines traditional sentiment analysis techniques with machine learning algorithms. Specifically, they employ sentiment dictionary extraction to identify sentiment-laden words, Support Vector Machines (SVM) for classifying sentiment polarity (positive, negative, or neutral), and Latent Dirichlet Allocation (LDA) for extracting sentiment topics within the reviews. This combined approach achieved high accuracy (80.36%) in sentiment analysis, providing valuable insights into E-commerce product experience.

Jeyapriya & Selvi (in [3]) propose a system for aspect-based opinion mining in customer reviews. It extracts product features (aspects) and their sentiment (positive/negative) using frequent itemset mining and a Naive Bayes classifier. Their system achieves high accuracy for both aspect extraction and sentiment orientation.

Alasmari (in [4]) analyzes massage roller reviews on Amazon to understand customer sentiment (positive/negative) and usage (offices, trips) using sentiment analysis and text mining (Python, Tableau). This helps product managers improve products based on customer feedback.

4 Methodology

4.1 Sentiment Analysis

This phase centers on evaluating the overall sentiment conveyed in textual data. The process entails a series of preprocessing steps, including converting text to lowercase, expanding contractions, handling negations, removing non-alphanumeric characters and extra spaces, tokenization, eliminating stopwords, lemmatization, and filtering out single-character to5ns and numerical values. Sentiment analysis techniques are applied to categorize text into positive, negative, or neutral sentiments based on the general tone of the text.

4.1.1. Text Preprocessing 11

Text preprocessing involves standardizing the text data to ensure consistency and accuracy in subsequent analysis. Techniques such as converting text to lowercase, expanding contractions, and handling negations help prepare the text for sentiment analysis tasks.

4.1.2. Sentiment Classification

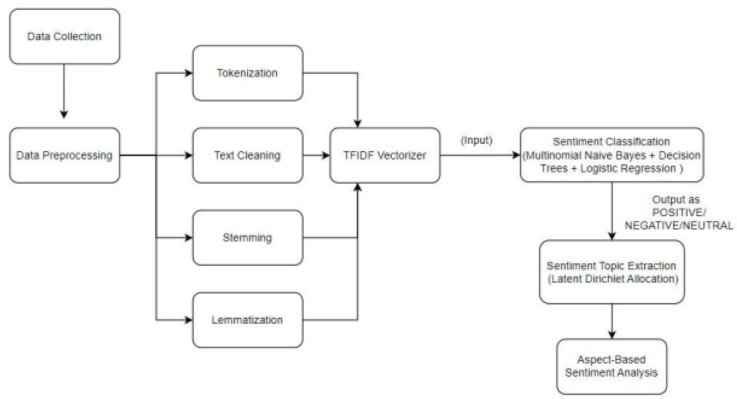
Sentiment classification techniques are utilized to assign sentiment labels (positive, negative, or neutral) to the text data based on the sentiment expressed within the text. This step is crucial for understanding the overall sentiment orientation of the text.

The following is formula for Logistic Regression:

$$ln\left(\frac{P}{1-P}\right) = a + bX$$

The sentiment analysis model combines multiple classifiers, including Multinomial Naive Bayes, Decision Trees, and Logistic Regression, to classify sentiments as positive, negative, or neutral. This ensemble approach leverages the strengths of each classifier to achieve high accuracy in sentiment classification tasks.

[Fig 1] Sentiment Analysis Architecture



4.2 Aspect-Based Sentiment Analysis (ABSA)

In this phase, the focus transitions to extracting specific aspects or features from the text and evaluating the sentiment associated with each aspect. Rule-based aspect extraction methods, Latent Dirichlet Allocation (LDA) topic modeling, and the integration of various aspect extraction techniques are employed to identify and assess aspects that influence sentiment expressions within the text.

4.2.1. Aspect Extraction Techniques

Aspect extraction techniques involve methodologies for identifying and extracting key aspects or features from the text data. These techniques may include rule-based approaches, dependency parsing, and other linguistic patterns to capture relevant aspects.

4.2.2. Latent Dirichlet Allocation (LDA) Analysis for Aspect Identification

LDA analysis is utilized to uncover latent topics and identify aspects within the text data. By modeling topics based on word co-occurrences, LDA aids in identifying key aspects that contribute to the overall sentiment expressed in the text.

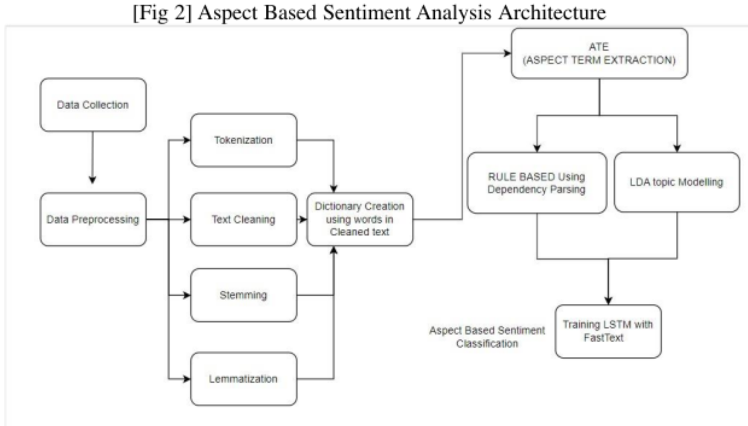
4.2.3. Integration of Aspect Extraction Methods

The integration of different aspect extraction methods involves combining rule-based extraction approaches with probabilistic models like LDA to provide a comprehensive view of the aspects present in the text data. By integrating these methods, a more holistic understanding of the text's aspects and sentiments can be achieved.

4.2.4. FastText Model Training for Aspect Similarity Calculations

The FastText model, known for its efficient word embedding capabilities and semantic similarity calculations, can enhance aspect identification and similarity measurements in aspect-based sentiment analysis tasks. By leveraging the FastText model, the analysis process benefits from the model's ability to capture word relationships and semantic meanings, aiding in accurate aspect identification and sentiment analysis.

[Fig 2] Aspect Based Sentiment Analysis Architecture



FastText model training focuses on capturing semantic relationships between words and aspects. By training the FastText model on preprocessed text data, the model can compute similarity scores between sentences and specified aspects, providing insights into the similarity and relevance of aspects identified in the text.

4.2.5 Aspect Sentiment classification using LSTM

The custom LSTM-based model used in this research project leverages the power of Long Short-Term Memory (LSTM) neural networks for text classification tasks. LSTM networks are a type of recurrent neural network (RNN) architecture known for their ability to capture long-term dependencies in sequential data. This makes LSTMs particularly well-suited for processing and analyzing text data, where understanding the context and relationships between words is crucial for accurate classification.

The LSTM model architecture consists of layers that incorporate memory cells, input gates, output gates, and forget gates, allowing the model to retain and update information over time. This architecture enables the LSTM model to effectively learn patterns and relationships within the text data, making it a popular choice for natural language processing tasks like sentiment analysis and text classification.

By utilizing a custom LSTM-based model, this research project aims to leverage the strengths of LSTM networks in capturing the sequential nature of textual data, extracting meaningful features, and predicting sentiment associated with identified aspects. The model's ability to learn from and remember long-range dependencies in text sequences is essential for accurately classifying sentiment and providing valuable insights into aspect-based sentiment analysis tasks.

5 Validations

5.1. Validation of Sentiment Analysis

5.1.1 Cross-Validation

Cross-validation is a common technique used to validate sentiment analysis models. K-fold cross-validation involves splitting the dataset into k subsets, training the model on k-1 subsets, and evaluating it on the remaining subset. This process is repeated k times, with each subset used as the validation set once. By averaging the performance metrics across the k folds, the model's generalization ability and performance can be assessed.

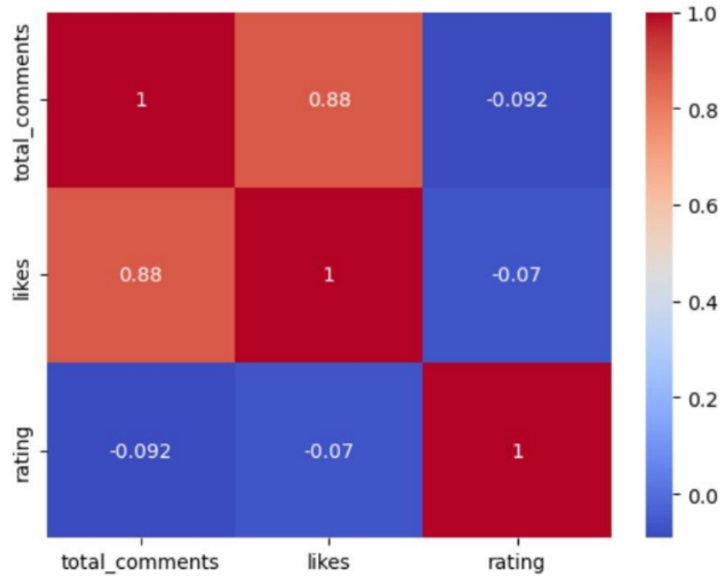
5.1.2 Performance Metrics

Performance metrics such as accuracy, precision, recall, and F1 score are commonly used to evaluate sentiment analysis models. Accuracy measures the overall correctness of the model's predictions. Precision and recall provide insights into the model's ability to correctly classify positive, negative, and neutral sentiments. The F1 score balances precision and recall, offering a combined measure of the model's performance.

5.1.3 Confusion Matrix Analysis

Analyzing the confusion matrix can reveal the model's performance in classifying sentiments. The matrix displays the true positive, true negative, false positive, and false negative predictions, allowing for a detailed examination of the model's strengths and weaknesses in sentiment classification.

[Fig 4] Confusion Matrix



5.2. Validation of Aspect-Based Sentiment Analysis:

5.2.1 Aspect Identification Accuracy

To validate aspect-based sentiment analysis, the accuracy of aspect identification is crucial. The model's ability to correctly extract and identify key aspects from the text data can be evaluated through precision, recall, and F1 score metrics specific to aspect extraction tasks.

5.2.2 Aspect-Sentiment Association

Validating the association between aspects and sentiments is essential in aspect-based sentiment analysis. The model's capability to correctly link aspects with the corresponding sentiments can be assessed through manual evaluation or automated measures such as aspect-sentiment coherence scores.

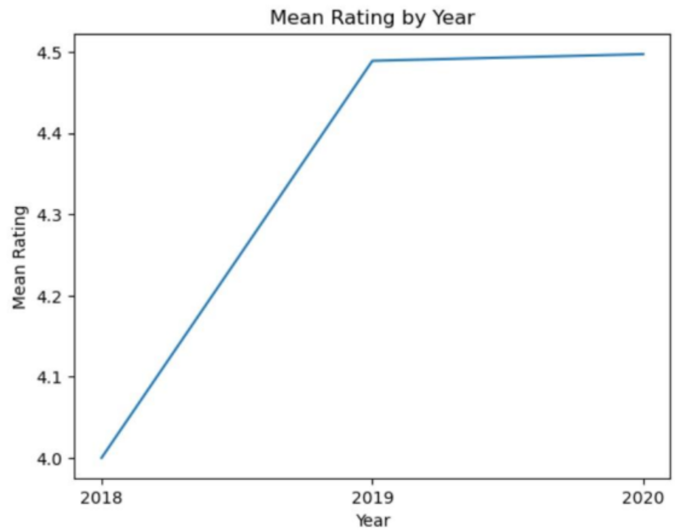
5.2.3 Cross-Domain Validation

Cross-domain validation involves testing the aspect-based sentiment analysis model on different domains or datasets to assess its generalizability and robustness. Evaluating the model's performance across diverse domains helps ensure that the aspect extraction and sentiment analysis techniques are applicable in various contexts.

5.2.4 Model Interpretability

The interpretability of the aspect-based sentiment analysis model is crucial for understanding the reasoning behind the model's predictions. Techniques such as feature importance analysis, attention mechanisms, and visualization tools can help interpret the model's decisions and provide insights into the aspects influencing sentiment.

[Fig 5] Mean Rating Analysis over the years



5.2.5 Conclusion

Validation of sentiment analysis and aspect-based sentiment analysis models is essential to ensure their accuracy, reliability, and applicability in real-world scenarios. By employing rigorous validation techniques and performance metrics, researchers can assess the models' effectiveness in capturing sentiment nuances and extracting key aspects from textual data. A well-validated sentiment analysis model can provide valuable insights for decision-making and sentiment analysis tasks in various domains.

6 Results

```
Naive Bayes
Accuracy Score: 0.8902245738948087
              precision    recall  f1-score   support

    negative       0.80      0.68      0.73      5646
     neutral       0.46      0.07      0.12      2048
    positive       0.91      0.98      0.94     33317

    accuracy                           0.89     41011
   macro avg       0.72      0.57      0.60     41011
weighted avg       0.87      0.89      0.87     41011
```

```
Logistic Regression
Accuracy Score: 0.9172417156372681
              precision    recall  f1-score   support

    negative       0.83      0.73      0.78      5646
     neutral       0.68      0.47      0.56      2048
    positive       0.94      0.98      0.96     33317

    accuracy                           0.92     41011
   macro avg       0.81      0.73      0.76     41011
weighted avg       0.91      0.92      0.91     41011
```

```
Decision Tree
Accuracy Score: 0.860086318304845
              precision    recall  f1-score   support

    negative       0.87      0.39      0.54      5646
     neutral       0.00      0.00      0.00      2048
    positive       0.86      0.99      0.92     33317

    accuracy                           0.86     41011
   macro avg       0.58      0.46      0.49     41011
weighted avg       0.82      0.86      0.82     41011
```

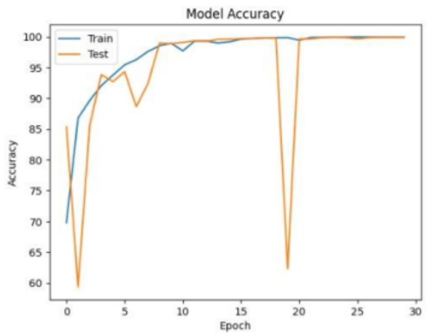
The results obtained from Naïve Bayes, Logistic Regression and Decision Trees is shown above. Out of the three we can see that **Logistic regression** is performing well with f1 score of **91%**

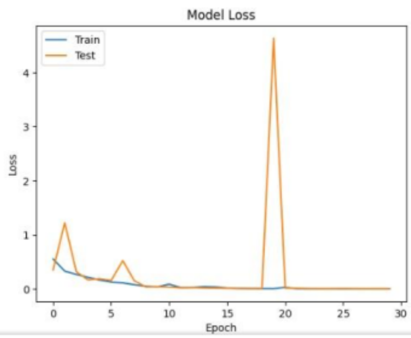
LDA for Topic Modelling:

```
[(0,
  '0.015*"phone" + 0.013*"camera" + 0.012*"go" + 0.011*"use" + 0.011*"battery" '
  '+ 0.010*"like" + 0.010*"get" + 0.009*"amazon" + 0.008*"day" + '
  '0.008*"samsung"'),
 (1,
  '0.038*"phone" + 0.021*"camera" + 0.020*"good" + 0.009*"samsung" + '
  '0.008*"screen" + 0.007*"get" + 0.007*"quality" + 0.007*"well" + '
  '0.006*"display" + 0.006*"like"'),
 (2,
  '0.039*"phone" + 0.017*"good" + 0.017*"camera" + 0.012*"use" + '
  '0.011*"battery" + 0.010*"get" + 0.010*"quality" + 0.009*"price" + '
  '0.008*"product" + 0.008*"time"'),
 (3,
  '0.030*"phone" + 0.021*"amazon" + 0.014*"day" + 0.013*"product" + '
  '0.012*"charge" + 0.012*"good" + 0.012*"dont" + 0.011*"work" + 0.010*"get" + '
  '0.010*"use"'),
 (4,
  '0.035*"phone" + 0.017*"buy" + 0.015*"good" + 0.014*"issue" + 0.013*"amazon" '
  '+ 0.012*"product" + 0.012*"charge" + 0.010*"dont" + 0.008*"mobile" + '
  '0.008*"face"'),
 (5,
  '0.018*"phone" + 0.015*"samsung" + 0.014*"mobile" + 0.012*"use" + '
  '0.012*"camera" + 0.012*"battery" + 0.011*"good" + 0.010*"quality" + '
  '0.009*"galaxy" + 0.007*"well"'),
 (6,
  '0.042*"phone" + 0.016*"good" + 0.013*"use" + 0.011*"battery" + 0.010*"get" '
  '+ 0.010*"camera" + 0.009*"one" + 0.009*"day" + 0.009*"charge" + '
  '0.008*"like"'),
 (7,
  '0.027*"phone" + 0.017*"camera" + 0.016*"good" + 0.014*"battery" + '
  '0.012*"mobile" + 0.011*"use" + 0.010*"charge" + 0.008*"samsung" + '
  '0.008*"product" + 0.008*"amazon"'),
 (8,
  '0.038*"phone" + 0.015*"good" + 0.014*"samsung" + 0.012*"camera" + '
  '0.012*"battery" + 0.011*"quality" + 0.010*"like" + 0.010*"mobile" + '
  '0.008*"one" + 0.007*"use"'),
 (9,
  '0.023*"phone" + 0.015*"good" + 0.012*"amazon" + 0.011*"battery" + '
  '0.011*"one" + 0.011*"samsung" + 0.010*"product" + 0.010*"camera" + '
  '0.010*"use" + 0.009*"dont"')]
```

Metrics from LSTM Training:

```
| Train Loss : 0.0016415671421571947 |  Val Loss : 0.0033114941863942628 | Train Acc : 99.93477630615234 | Val Acc : 99.88401794433594 |
```





Aspect Term Extraction and Aspect Based Sentiment Analysis:

```
I am really impressed with the phone's battery backup.
Aspect :  battery
Sentiment :  Positive
```

7 Discussion

In the realm of sentiment analysis, the transition from the initial phase focusing on overall sentiment to aspect-based sentiment analysis (ABSA) represents a crucial evolution in understanding textual data's nuanced aspects. ABSA delves deeper into the finer details of text, extracting specific features or aspects and evaluating the sentiment associated with each. While traditional sentiment analysis categorizes text into broad positive, negative, or neutral sentiments, ABSA offers a more granular approach, dissecting the text to discern sentiments tied to individual aspects or topics. This shift is paramount, especially in domains where nuanced opinions about various aspects of a product, service, or experience are pivotal, such as customer reviews, social media sentiment, or market research analysis. By employing techniques like rule-based aspect extraction and probabilistic modeling such as Latent Dirichlet Allocation (LDA), ABSA equips analysts with the tools to dissect complex texts, uncovering the sentiment nuances buried within.

Moreover, the integration of aspect extraction methods, including both rule-based approaches and probabilistic models like LDA, underscores the multidimensional nature of textual sentiment analysis. By combining diverse methodologies, analysts can obtain a more comprehensive understanding of the text's nuances and sentiment dynamics. This integration allows for a holistic exploration of the text's underlying aspects, from explicit features to latent topics, enabling a richer interpretation of sentiment expressions. Additionally, leveraging advanced models like FastText for aspect similarity calculations further enhances the ABSA process by capturing intricate semantic relationships between words and aspects. This sophisticated approach not only refines aspect identification but also provides insights into the relative relevance and similarity of identified aspects, empowering analysts to extract deeper insights from textual data and make more informed decisions based on nuanced sentiment analysis.

8 Conclusion

In conclusion, the journey through sentiment analysis, from the initial phase to aspect-based sentiment analysis (ABSA), underscores the progression towards a more nuanced understanding of textual data's sentiment dynamics. While sentiment analysis provides a broad overview of sentiment orientations, ABSA delves deeper, dissecting texts to uncover sentiment nuances associated with specific aspects or topics. By employing a combination of techniques such as rule-based aspect extraction, probabilistic modeling with LDA, and leveraging advanced models like FastText for aspect similarity calculations, ABSA equips analysts with powerful tools to extract deeper insights from textual data. This multidimensional approach enables a more comprehensive understanding of the sentiments expressed, facilitating richer interpretations and informed decision-making across various domains, from customer feedback analysis to market research and beyond. As the field continues to evolve, ABSA stands as a testament to the importance of granularity in sentiment analysis, enabling analysts to navigate the complexities of textual data with greater precision and depth.

9 Future Scope

In the future, sentiment analysis and aspect-based sentiment analysis (ABSA) are poised to witness significant advancements and widespread application. Innovations in machine learning algorithms and deep learning architectures will refine sentiment analysis models, enabling them to capture nuanced sentiments with greater accuracy and efficiency. Additionally, the integration of multimodal data sources and the development of scalable processing techniques will expand the scope of sentiment analysis to encompass diverse forms of multimedia content and handle large-scale datasets more effectively. Beyond traditional domains like marketing, sentiment analysis will find applications in fields such as healthcare, finance, and governance, offering valuable insights for decision-making processes. As research and technological progress continue to drive the evolution of sentiment analysis, its role in understanding human sentiment across various contexts will become increasingly vital.

plagarism_check.docx

ORIGINALITY REPORT

9%	6%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	2%
2	Ritu Chauhan, Anika Goel, Bhavya Alankar, Harleen Kaur. "Predictive modeling and web-based tool for cervical cancer risk assessment: A comparative study of machine learning models", MethodsX, 2024 Publication	1%
3	Submitted to Berlin School of Business and Innovation Student Paper	1%
4	Huaqian He, Guijun Zhou, Shuang Zhao. "Exploring E-commerce Product Experience based on Fusion Sentiment Analysis Method", IEEE Access, 2022 Publication	1%
5	"Data Science and Applications", Springer Science and Business Media LLC, 2024 Publication	1%
6	www.jmest.org Internet Source	1%

7 www.codinginterviewpro.com
Internet Source
1 %

---

8 Pokkuluri Kiran Sree, SSSN Usha Devi N, Phaneendra Varma Chintalapati, Gurujukota Ramesh Babu, PBV Raja Rao. "Chapter 11 Drug Recommendations Using a Reviews and Sentiment Analysis by RNN", Springer Science and Business Media LLC, 2024
Publication
<1 %

---

9 AVOM Désiré, ONGO NKOA Bruno Emmanuel, ATANGANA ZAMBO Charles Christian, KAMDOUM KAMWA Ulrich Kevin et al. "Colonial Origins and Growth of Financial Markets in Africa: A comparative analysis based on Institutions", Journal of Government and Economics, 2024
Publication
<1 %

---

10 Ying Zhang. "Artificial intelligence carbon neutrality strategy in sports event management based on STIRPAT-GRU and transfer learning", Frontiers in Ecology and Evolution, 2023
Publication
<1 %

---

11 Submitted to University of Sunderland
Student Paper
<1 %

---

12 Submitted to Deakin University
Student Paper
<1 %

13	ej-eng.org Internet Source	<1 %
14	Ziheng Cui, Shuai Huang, Meng Wang, Kaili Nie, Yunming Fang, Tianwei Tan. "Improving the CFPP property of biodiesel via composition design: An intelligent raw material selection strategy based on different machine learning algorithms", Renewable Energy, 2021 Publication	<1 %
15	RIMPY DAHIYA, Amita Dhankhar, AKSHIKA DHANKHAR. "Sentimental Analysis Of Social Networks :A Comprehensive review(2018-2023)", Research Square Platform LLC, 2023 Publication	<1 %
16	doi.org Internet Source	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		