International Journal of Applied Engineering & Technology

ASPECT-BASED SENTIMENT ANALYSIS: A METHODOLOGICAL FRAMEWORK FOR PRECISE SENTIMENT IDENTIFICATION IN TEXTUAL DATA

Srawan Nath[1], Richa Rawal[2] and Anil Pal[3]
[1,2]Research Scholar and [3]Associate Professor, Computer Science and Engg., Suresh Gyan Vihar University Jaipur
[1]nath.sarwan@gmail.com

ABSTRACT
This paper introduces a methodological framework for aspect-based sentiment analysis aimed at achieving precise sentiment identification in textual data. The study focuses on developing an effective approach for analysing sentiments towards specific aspects within text, considering the challenges posed by the high dimensionality and sparsity of data, which often complicate clustering in unsupervised learning scenarios. Latent Dirichlet Allocation (LDA) is a prominent unsupervised machine learning algorithm extensively used in natural language processing (NLP) for topic modelling is investigated in conjunction with various clustering techniques on a dataset comprising product reviews of iPhones from Amazon.com. The framework systematically extracts aspects, performs sentiment classification, and aggregates results to discern nuanced sentiments. The study demonstrates the effectiveness of this approach in accurately capturing sentiments related to diverse aspects discussed in textual data. The framework's novelty lies in its comprehensive integration of aspect identification and sentiment analysis, offering a robust solution for understanding nuanced sentiment expressions in text data. Additionally, the paper explores the application of clustering techniques for sentiment analysis on the dataset of iPhone reviews, employing algorithms such as K-means, hierarchical clustering, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to categorize reviews into coherent groups based on similarity in content or sentiment. Evaluation of clustering performance metrics, including silhouette scores, is conducted to assess the effectiveness of each method in partitioning the review data. Extensive experiments are designed, and the results demonstrate the effectiveness and high silhouette score of the LDA–clustering approach.

Keywords: LDA, Sentiment analysis, iPhone reviews, Clustering techniques, Silhouette score

1. INTRODUCTION
Aspect-Based Sentiment Analysis (ABSA) has emerged as a significant area of research in sentiment analysis, aiming to dissect textual data into granular aspects and accurately discern sentiments toward specific elements. Unlike traditional sentiment analysis methods that focus on overall sentiment without considering specific aspects, ABSA seeks to capture nuanced opinions and attitudes towards individual features or entities discussed within the text. Despite its importance in domains such as product reviews and social media discussions, ABSA faces challenges in aspect extraction, sentiment classification, and aspect-level sentiment aggregation [1]. Existing methods often rely on manual feature engineering or domain-specific rules, limiting scalability and adaptability across different domains and languages. The inherent ambiguity and complexity of natural language further complicate accurate aspect and sentiment identification, particularly in informal text sources like social media [2].

The novelty of this topic lies in its comprehensive integration of aspect-based sentiment analysis techniques to achieve precise sentiment identification within textual data. By addressing challenges in aspect extraction, sentiment classification, and aspect-level sentiment aggregation, our methodological framework offers a robust solution for capturing nuanced sentiments towards specific elements discussed in text, contributing to advancements in sentiment analysis methodologies and providing valuable insights across various domains and applications [4]. In the contemporary digital landscape, customer reviews serve as invaluable sources of information for businesses aiming to understand consumer sentiments and preferences [5]. Among the myriad of products scrutinized by consumers, the iPhone stands as a quintessential example, with its releases eagerly anticipated and extensively reviewed. Analysing iPhone reviews offers businesses a glimpse into the perceptions, satisfaction levels, and areas for improvement as perceived by consumers [6]. In this study, we delve into the realm of sentiment analysis by employing clustering techniques to dissect and categorize iPhone reviews.

Through the application of clustering algorithms such as K-means, hierarchical clustering, and DBSCAN, we aim to uncover underlying patterns and themes within the reviews, enabling a nuanced understanding of customer sentiments [7]. By evaluating the performance of these clustering methods and interpreting the resulting clusters, we seek to extract actionable insights that can inform product development strategies, marketing campaigns, and customer engagement initiatives [8]. Through this analysis, we endeavor to shed light on the multifaceted landscape of iPhone reviews and illuminate pathways for businesses to enhance customer satisfaction and drive business success in the ever-evolving digital marketplace [9, 10].

In this context, this paper proposes a methodological framework for aspect-based sentiment analysis that addresses these challenges and provides a systematic approach for precise sentiment identification in textual data. By leveraging state-of-the-art unsupervised machine learning algorithms and different clustering methods, prposed framework aims to enhance the efficiency of aspect-based sentiment analysis across various text sources and domains. We demonstrate the effectiveness of our approach through empirical evaluations on diverse datasets, highlighting its ability to capture nuanced sentiments towards specific aspects and entities discussed in textual data. Through this research, we contribute to the advancement of sentiment analysis methodologies and provide valuable insights for applications such as product recommendation systems, opinion mining, and brand reputation management.

2. RELATED WORK

Previous sentiment analysis research has often focused on analysing overall sentiment in text, lacking granularity to capture nuanced opinions towards specific aspects. Aspect-based sentiment analysis (ABSA) addresses this by dissecting text into fine-grained aspects and identifying sentiment associated with each [1,5]. ABSA has been applied across domains like social media and customer feedback but still faces challenges in aspect extraction and sentiment classification. Our research proposes a methodological framework for ABSA leveraging NLP techniques and machine learning to enhance accuracy and efficiency. We introduce a novel neural network architecture, AS-GCN, achieving state-of-the-art performance in aspect-based sentiment classification by capturing semantic relations between words and aspects [6,7].

A hybrid neural network model combining convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) was introduced for aspect-based sentiment analysis. Incorporating attention mechanisms to dynamically weigh the importance of words and aspects improved sentiment classification accuracy [10]. Additionally, research explored integrating external knowledge sources like knowledge graphs to enrich aspect and sentiment representations, enhancing nuanced relationships between entities and sentiments. Domain adaptation techniques, such as adversarial training, were investigated to improve model generalization across different domains. Recent advancements, including a BERT-based model for aspect-sentiment extraction and a comparative study of deep learning architectures, contribute to ABSA methodologies, addressing challenges and facilitating applications across domains [11].

A gated convolutional network (GCN) architecture was introduced for aspect-based sentiment analysis, utilizing gated mechanisms to capture informative features from aspects and sentiments, enhancing sentiment classification performance. Another study proposed a BERT-based model with context attention mechanisms, achieving improved sentiment classification and aspect extraction by attending to relevant context [12]. Data augmentation combined with pre-trained language models addressed data scarcity in aspect-based sentiment analysis, leading to enhanced performance [13]. A joint aspect and sentiment classification framework based on BERT performed aspect extraction and sentiment classification simultaneously, leveraging contextualized representations for effective relationship capture. Contrastive learning enhanced discriminative aspect and sentiment representations, while a dual relation attention network improved sentiment classification by modelling aspect-sentiment interactions from multiple perspectives [14].

A hierarchical transformer model was proposed for aspect-based sentiment analysis, utilizing hierarchical attention mechanisms to capture interactions from words to aspects, enhancing sentiment classification at the aspect level [15]. Another study introduced a graph neural network (GNN) approach, effectively capturing contextual information and semantic relationships between aspects and sentiments by modelling dependencies as a graph structure, leading to improved sentiment classification accuracy. Investigating multi-task learning techniques, researchers jointly optimized aspect extraction and sentiment classification tasks, improving generalization ability and performance on aspect-based sentiment analysis tasks [16]. An adversarial domain adaptation approach addressed domain shift by aligning feature distributions between domains using adversarial training, improving sentiment classification on target domain data [17]. Few-shot learning techniques were explored for aspect-based sentiment analysis, leveraging a small number of annotated samples per aspect to generalize to new aspects effectively. Lastly, an attention-guided graph convolutional network (AGGCN) utilized attention mechanisms to capture contextual information and semantic relationships between aspects and sentiments, leading to improved sentiment classification accuracy [18].

A meta-learning approach was proposed for cross-domain aspect-based sentiment analysis, effectively transferring knowledge from source domains to target domains with limited labeled data [19]. Transformer-based models, including Bidirectional Encoder Representations from Transformers (BERT) and robustly optimized BERT approach (RoBERTa), achieved state-of-the-art performance in sentiment classification and aspect extraction tasks, demonstrating their effectiveness for analysing sentiments towards specific aspects in textual data [20]. In the realm of sentiment analysis, clustering techniques offer a robust approach to distil large datasets into coherent groups, facilitating the identification of prevalent themes and sentiments. K-means clustering has been extensively applied to segment consumer reviews into clusters reflecting different aspects of the experience, while hierarchical clustering allows for a deeper exploration of nested clusters, providing insights into the hierarchical organization of sentiments within the dataset [21]. Density-based clustering algorithms like DBSCAN have shown effectiveness in handling irregular shapes and varying densities of clusters, particularly in categorizing social media comments into sentiment polarities. This study aims to contribute to sentiment analysis literature by leveraging clustering algorithms to analyse iPhone reviews, providing actionable insights for businesses to understand and respond to consumer sentiments effectively [22].

3. METHODOLOGICAL FRAMEWORK

The research methodological framework for aspect-based sentiment analysis involves several key steps aimed at developing and evaluating effective techniques for precise sentiment identification towards specific aspects in textual data. The typical research methodological framework is depicted in figure 1:

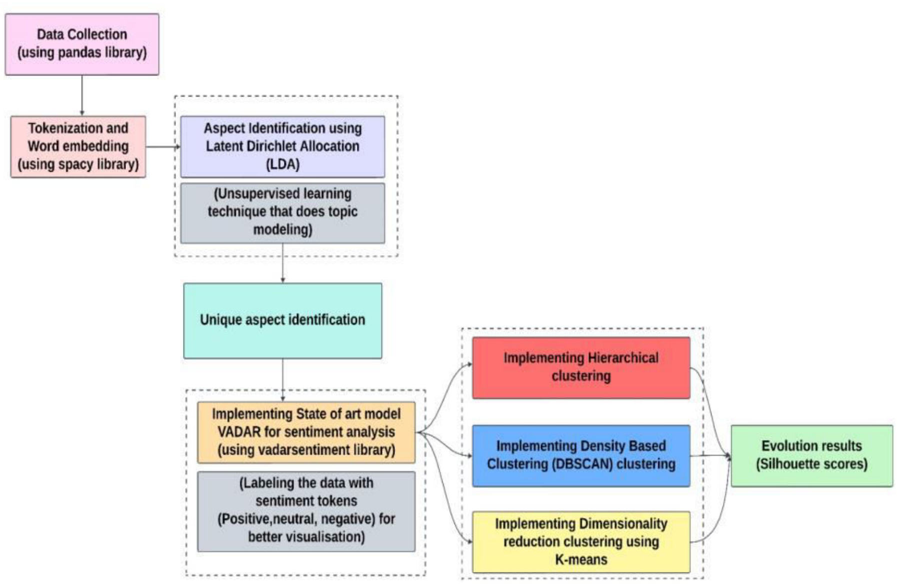

Figure 1: Proposed methodological framework

3.1 Dataset Collection

Dataset product review of iPhone on Amazon .com is collected for the further analysis of aspect-based specific opinion for Precise Sentiment Identification in Textual Data. Data Collection gather relevant textual data sources that contain information about the aspects of interest and associated sentiments as shown in figure 2.

	index	product	review_text	review_rating	review_title
0	0	Apple iPhone XR (64GB) - Black	NOTE:	3.0 out of 5 stars	Which iPhone you should Purchase ? iPhone 8, X...
1	1	Apple iPhone XR (64GB) - Black	Very bad experience with this iPhone xr phone....	1.0 out of 5 stars	Don't buy iPhone xr from Amazon.
2	2	Apple iPhone XR (64GB) - Black	Amazing phone with amazing camera coming from ...	5.0 out of 5 stars	Happy with the purchase
3	3	Apple iPhone XR (64GB) - Black	So I got the iPhone XR just today. The product...	1.0 out of 5 stars	Amazon is not an apple authorised reseller. Pl...
4	4	Apple iPhone XR (64GB) - Black	I've been an android user all my life until I ...	5.0 out of 5 stars	Excellent Battery life and buttery smooth UI

Figure 2: Dataset product review of iPhone

3.2 Tokenisation and Word Embedding

Aspect-based sentiment analysis (ABSA) requires effective tokenization and word embedding techniques to capture nuanced opinions expressed in text, especially in datasets related to products like the iPhone. Tokenization is the process of breaking down text into smaller units, typically words or subwords, known as tokens. For aspect-based sentiment analysis on iPhone-related datasets, tokenization involves segmenting the text into meaningful units such as words, phrases, or even specific aspects related to the iPhone (e.g., battery life, camera quality, user interface). Since aspect-based sentiment analysis focuses on extracting opinions about specific aspects, it's essential to tokenize the text in a way that preserves the aspect-related information.

Word embedding is a technique used to represent words as dense vectors in a high-dimensional space, where words with similar meanings are closer to each other. Pre-trained word embeddings such as doc2bow, Word2Vec, GloVe, or FastText can be used to capture semantic relationships between words in the context of iPhone-related text data. Aspect-specific word embeddings can be learned or adapted during the training process to better capture the nuances of opinions related to different aspects of the iPhone.

3.3 Aspect Identification using Latent Dirichlet Allocation

Aspect identification using Latent Dirichlet Allocation (LDA) involves preprocessing the text data by cleaning, tokenizing, and constructing a document-term matrix; then applying LDA to extract latent aspects. Through LDA, topics representing aspects are inferred from the corpus, where each topic is characterized by a distribution of terms. These topics are interpreted by examining the top words associated with each, and human-interpretable labels are assigned. Documents are then assigned to dominant topics based on their topic distributions. The resulting aspect identification can be used for various downstream tasks such as sentiment analysis or recommendation systems. Latent Dirichlet Allocation (LDA) is a prominent unsupervised machine learning algorithm extensively utilized in natural language processing (NLP) for topic modelling. Operating under the assumption that each document in a corpus is a mixture of latent topics, and each word in the document is attributable to one of these topics, LDA aims to infer the underlying thematic structure of the text data. Through a probabilistic generative process, LDA iteratively estimates the distributions of topics across documents and words within topics, maximizing the likelihood of observing the corpus given the model parameters. The resulting topic distributions offer insights into the main themes present in the documents, while the word distributions reveal the prominent terms associated with each topic. LDA finds applications in document clustering, topic summarization, information retrieval, and sentiment analysis, contributing significantly to the analysis and understanding of large text collections in machine learning and NLP tasks.

3.4 Implementing State of art Model VADER for Sentiment Analysis

Implementing VADER for sentiment analysis involves first importing the Sentiment Intensity Analyzer class from the vader Sentiment library in Python. After initializing a Sentiment Intensity Analyzer object, which serves as the sentiment analysis tool, the text data to be analysed is provided as input. By calling the polarity_scores() method of the analyser object with the text, sentiment scores are generated, including the compound score representing the overall sentiment as presented in figure 3. This approach makes VADER particularly useful for quick sentiment analysis tasks, especially in social media or informal text contexts where it has demonstrated effectiveness. However, it may not be as suitable for longer or more formal texts compared to machine learning-based approaches. Nevertheless, its ease of use and pre-trained model make it a valuable tool for sentiment analysis tasks, especially when training data is limited or unavailable. It is used for labelling the data with sentiment tokens for better visualization.

	index	product	review_text	review_rating	review_title	tokens	dominant_topic	topic_name	sentiment_vader
0	0	Apple iPhone XR (64GB) - Black	NOTE:	3.0 out of 5 stars	Which iPhone you should Purchase ? iPhone 8, X...	[note]	1	Buying Experience	Neutral
1	1	Apple iPhone XR (64GB) - Black	Very bad experience with this iPhone xr phone....	1.0 out of 5 stars	Don't buy iPhone xr from Amazon.	[bad, experience, iPhone, xr, phone, camera, f...	2	Product Review	Negative
2	2	Apple iPhone XR (64GB) - Black	Amazing phone with amazing camera coming from ...	5.0 out of 5 stars	Happy with the purchase	[amazing, phone, amazing, camera, come, iPhone...	1	Buying Experience	Positive
3	3	Apple iPhone XR (64GB) - Black	So I got the iPhone XR just today. The product...	1.0 out of 5 stars	Amazon is not an apple authorised reseller. Pl...	[get, iPhone, xr, today, product, look, amazin...	1	Buying Experience	Positive
4	4	Apple iPhone XR (64GB) - Black	I've been an android user all my life until I ...	5.0 out of 5 stars	Excellent Battery life and buttery smooth UI	[android, user, life, decide, try, iPhone, XR,...	3	Customer Satisfaction	Positive

Figure 3: Refined text data to be analysed is provided as input

3.5 Clustering

In this framework, three distinct clustering methodologies are employed: hierarchical clustering, DBSCAN clustering, and dimensionality reduction clustering via K-means. The evaluation of the clustering performance concerning aspect-specific sentiment opinions is gauged through the Silhouette score, a metric that assesses the quality of clustering.

3.6 Evaluation

Silhouette score is determined of each clustering technique to identified the performance. The Silhouette score is a metric used to evaluate the quality of clustering in unsupervised machine learning. It measures how similar an object is to its own cluster compared to other clusters. The significance of the Silhouette score lies in its ability to provide a quantitative measure of the compactness and separation of clusters, helping to identify the optimal number of clusters and assess the overall effectiveness of the clustering algorithm. A high Silhouette score (close to +1) indicates that the data point is well-clustered, with instances within the same cluster being close to each other and well-separated from instances in other clusters. On the other hand, a low Silhouette score (close to -1) suggests that the data point may have been assigned to the wrong cluster. Therefore, a higher Silhouette score generally indicates better clustering performance.

4. RESULTS AND DISCUSSION

Clustering holds significant importance in sentiment analysis as it enables the organization and interpretation of sentiment-bearing text data. Through clustering, similar sentiments expressed across various texts can be grouped together, aiding in the identification of different aspects or themes under discussion and the sentiments associated with them. This process facilitates a deeper understanding of opinion dynamics within the dataset, allowing analysts to discern how sentiments evolve over time or in response to specific topics or events. Moreover, clustering serves as a means of feature extraction, extracting meaningful representations from text data for building more effective sentiment classification models or summarizing large volumes of data. Additionally, clustering enables the segmentation of users or customers based on their sentiment profiles, paving the way for personalized recommendations or targeted interventions. Furthermore, clustering assists in evaluating and validating sentiment analysis models by comparing the generated clusters with human-labeled data, thereby assessing model performance and identifying areas for improvement. Overall, clustering plays a vital role in organizing, summarizing, and understanding sentiment within text data, ultimately enhancing sentiment analysis and decision-making processes.

4.1 Distribution of Dominant Topics

The distribution of dominant topics refers to the spread or prevalence of various topics across a corpus of text data. In the context of topic modelling, particularly using techniques like Latent Dirichlet Allocation (LDA), each document is typically associated with one or more topics, with one topic being dominant. Analysing the distribution of dominant topics involves examining the frequency or proportion of different topics present in the dataset as depicted in figure 4. For example, if we have performed topic modelling on a collection of customer reviews for a product, the distribution of dominant topics would indicate which aspects or themes of the product are most frequently discussed. This analysis could reveal insights such as whether customers primarily discuss the product's features, performance, pricing, or customer service. Understanding the distribution of dominant topics can provide valuable insights into the overall focus or sentiment of the text data. It can help identify prevalent themes, trends, or areas of concern, guiding further analysis or decision-making processes. Additionally, comparing the distribution of dominant topics across different datasets or time periods can reveal changes in customer preferences, emerging trends, or the effectiveness of interventions.

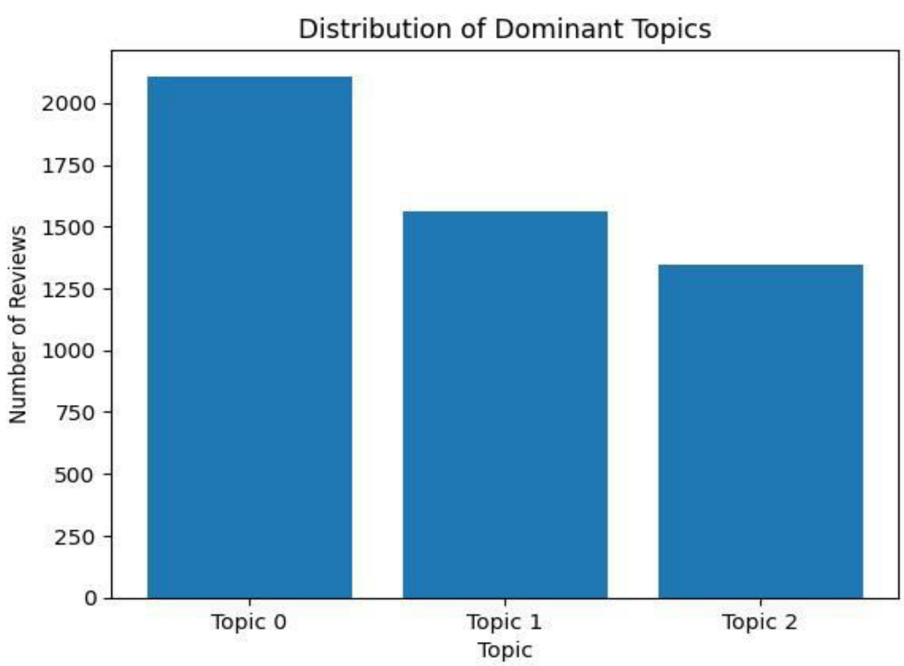

Figure 4: Distribution of dominant topics in the dataset

4.2 Dendrogram

A dendrogram is a graphical representation of hierarchical clustering results. In the context of sentiment analysis, a dendrogram can be used to visualize the hierarchical structure of clusters formed based on similarity in sentiment aspects or opinions expressed in text data. Each leaf node in the dendrogram represents a document or a data point, and the branches illustrate the clustering process, showing how documents are grouped together based on their similarity. Dendrograms are particularly useful for understanding the relationships between different clusters and identifying the level of granularity at which clusters are formed. They provide insights into the hierarchical organization of sentiment aspects or opinions, allowing analysts to explore the underlying structure of the data and make informed decisions about the number of clusters to consider. By visualizing the dendrogram, analysts can identify clusters of documents that exhibit similar sentiment patterns and explore the hierarchy of sentiment aspects present in the data. This can aid in gaining a deeper understanding of the nuanced relationships between different sentiment expressions and in identifying key themes or topics that are prevalent across the dataset. Additionally, dendrograms can help in evaluating the quality of clustering results and in identifying outliers or anomalies that may require further investigation.

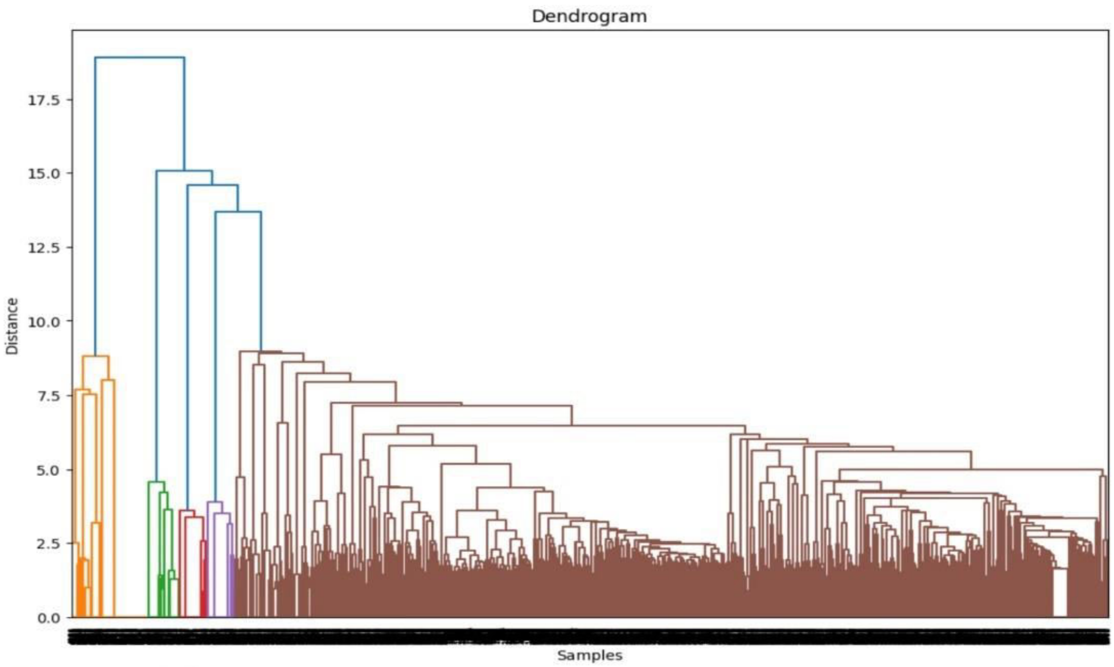

Figure 5: Graphical representation of hierarchical clustering

4.3 Density-Based Spatial Clustering of Applications with Noise

Clustering iPhone reviews using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm involves first preparing and preprocessing the review data, including steps such as text cleaning and feature extraction. Next, the DBSCAN algorithm is applied to the feature vectors derived from the pre-processed text data, with careful consideration given to setting parameters such as eps (epsilon) and min_samples. Through this process, DBSCAN automatically identifies clusters of reviews based on their proximity in the feature space, while also labeling some points as noise (outliers). Following clustering, the quality of the results is evaluated using metrics like silhouette score, and the clusters are interpreted to discern common themes or sentiments present in the iPhone reviews. Additionally, visualization techniques may be employed to gain further insights into the distribution of reviews in the feature space and the clustering structure as illustrated in figure 6. Overall, this approach facilitates the identification of key aspects and sentiments expressed in iPhone reviews, aiding in understanding customer opinions and preferences.

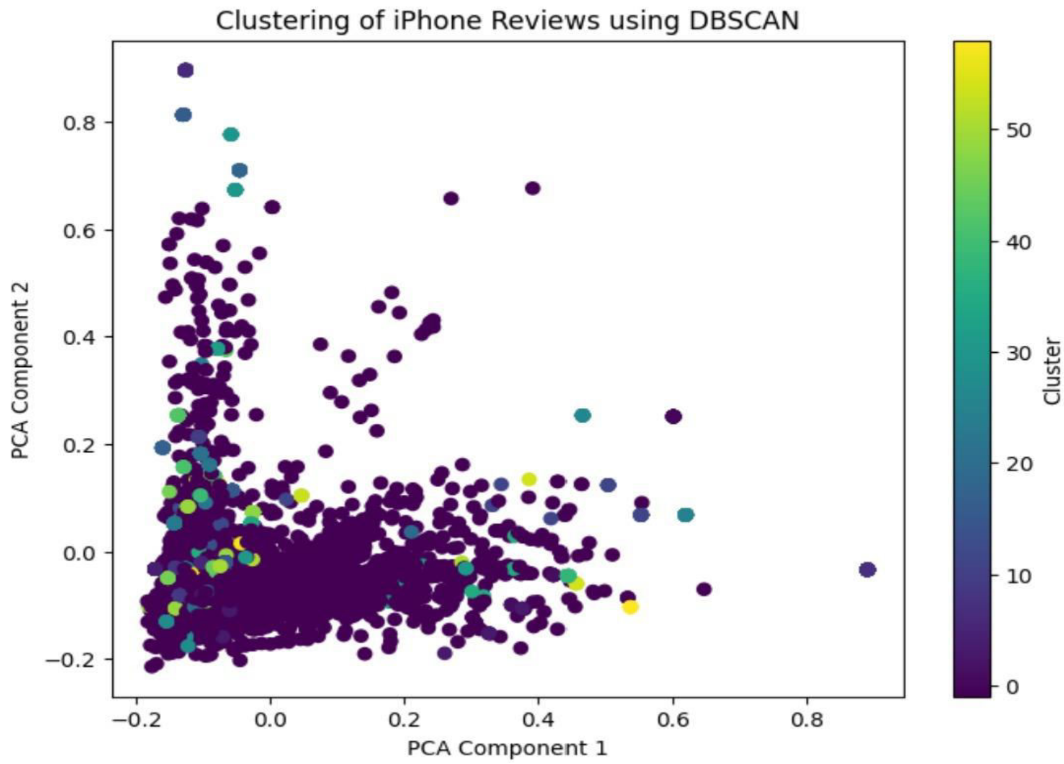

Figure 6: Graphical representation of DBSCAN clustering

4.4 Clustering iPhone Reviews using Dimensionality Reduction and K-means

Clustering iPhone reviews using dimensionality reduction and K-means entails preprocessing the text data and converting it into numerical feature vectors is showing in figure 7. Techniques such as TF-IDF or word embeddings are commonly employed for this purpose. Following feature extraction, dimensionality reduction methods like Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbour Embedding (t-SNE) are applied to reduce the dimensionality of the feature space while preserving essential information. Subsequently, the K-means algorithm is employed to identify clusters of similar reviews based on the reduced-dimensional feature vectors. By iteratively assigning data points to cluster centroids and updating centroids until convergence, K-means produces clusters representative of distinct sentiments or aspects expressed in the reviews. Evaluation metrics such as silhouette score are utilized to assess the quality of clustering results, guiding adjustments in parameters and preprocessing techniques as needed. The interpretation of resulting clusters allows for the identification of prevalent themes and sentiments across iPhone reviews, aiding in understanding customer opinions and preferences regarding different aspects of the product. Additionally, visualization techniques may be employed to visually explore the clustering structure and distribution of reviews in the reduced-dimensional space, facilitating further insights into customer sentiments.

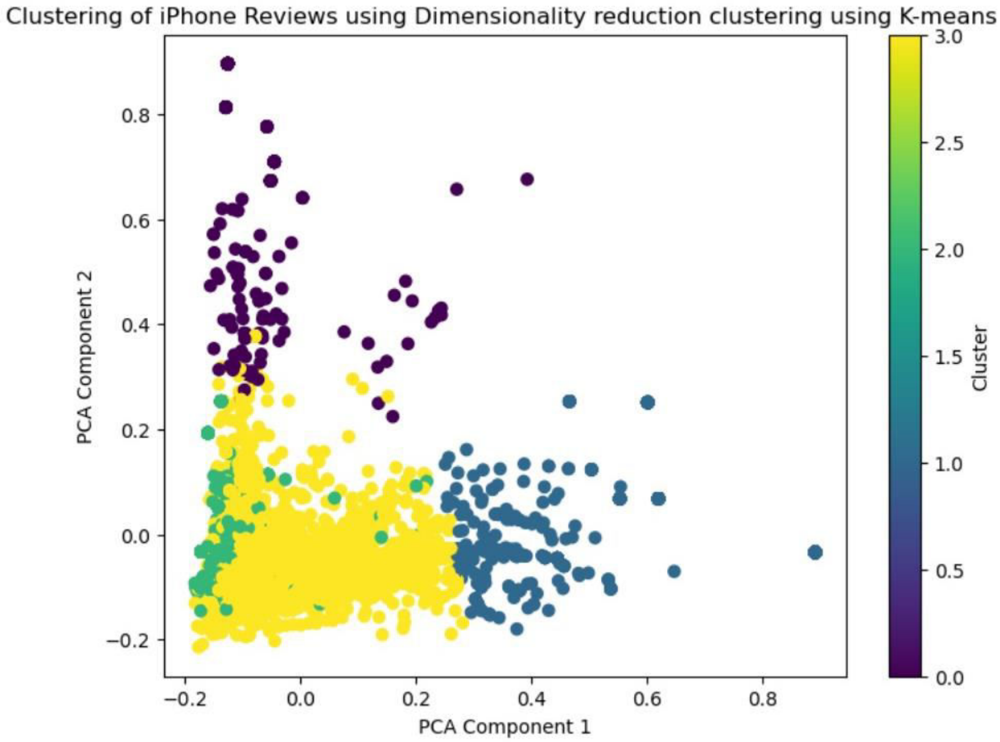

Figure 7: Graphical representation of dimensionality reduction and K-means

4.5 Elbow Method

The elbow method serves as a heuristic approach to determine the optimal number of clusters for clustering algorithms like K-means. In the context of analysing iPhone reviews, this method involves preprocessing the textual data and transforming it into numerical feature vectors. Subsequently, K-means clustering is applied across a range of cluster numbers, with the within-cluster sum of squares (WCSS) computed for each configuration. By plotting the number of clusters against their corresponding WCSS values, an elbow point on the plot is identified and shown in figure 8, indicating a significant decrease in the rate of WCSS reduction. This elbow point signifies the optimal number of clusters, striking a balance between maximizing cluster separation and minimizing cluster count. Choosing the number of clusters at this point enables effective segmentation of iPhone reviews into distinct themes or sentiments expressed by customers. Thus, leveraging the elbow method facilitates more meaningful interpretation of customer feedback, offering valuable insights into various aspects of the iPhone product and enhancing decision-making processes.

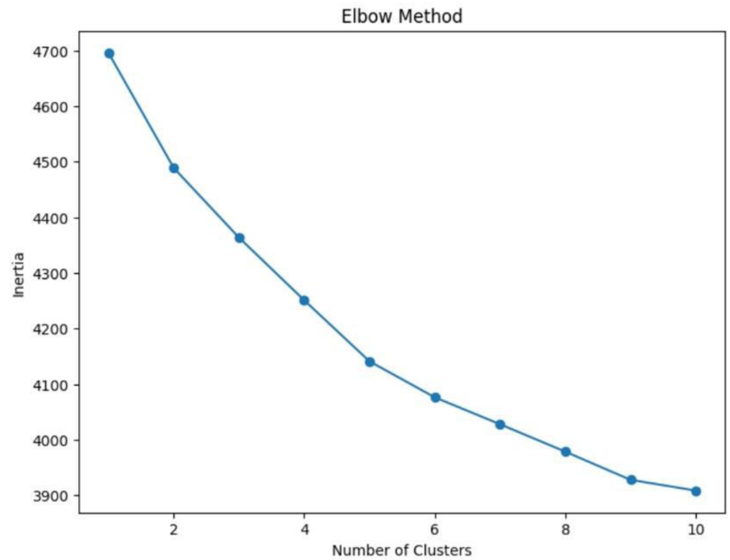

Figure 8: elbow method for clustering algorithms like K-means

4.6 Histogram of Cluster Sizes

Creating a histogram of cluster sizes involves visualizing the distribution of data points across clusters after applying the K-means clustering algorithm to iPhone reviews. Initially, K-means clusters the pre-processed reviews into a predetermined number of clusters. Subsequently, the number of data points assigned to each cluster is counted. By plotting this count against the respective cluster sizes, the resulting histogram illustrates the frequency distribution of cluster sizes as given in figure 9. Analysing the histogram provides insights into the variability of cluster sizes and the homogeneity or heterogeneity of the formed clusters. For instance, a skewed histogram with a few large clusters and several smaller ones may indicate an imbalance in the distribution of sentiments or themes within the reviews. Conversely, a histogram with evenly distributed cluster sizes suggests a more balanced representation of sentiments or topics. Thus, generating a histogram of cluster sizes facilitates a deeper understanding of the clustering outcome and aids in interpreting the clustering results in the context of iPhone reviews.

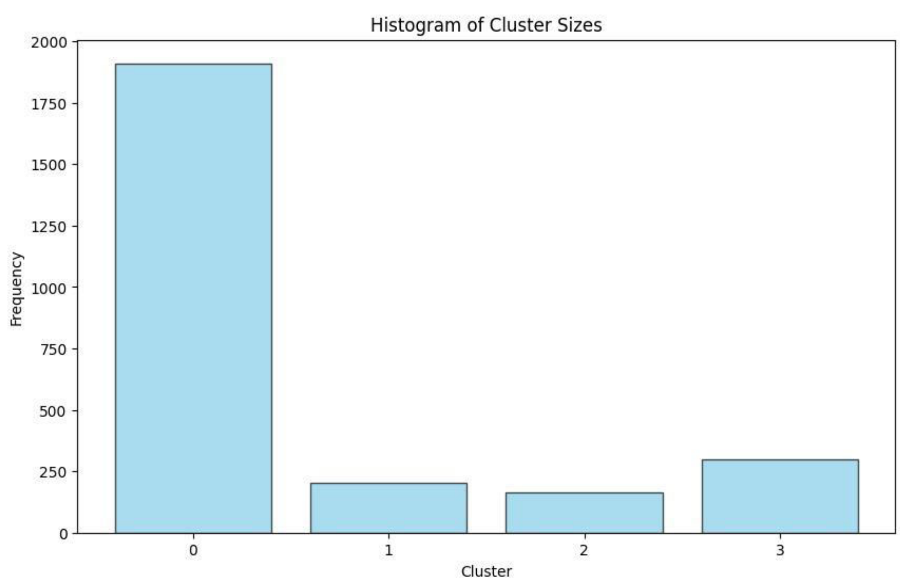

Figure 9: Histogram of cluster sizes vs. frequency

4.7 Comparing Silhouette Scores across Different Clustering Methods

Comparing silhouette scores across different clustering methods serves as a crucial step in determining the efficacy of each approach in segmenting the iPhone reviews dataset into meaningful clusters. Firstly, for K-means clustering, the algorithm is applied with varying numbers of clusters, and silhouette scores are computed for each configuration. The optimal number of clusters is then selected based on the highest silhouette score attained. Secondly, hierarchical clustering techniques such as agglomerative clustering are employed, with silhouette scores computed across different linkage methods. The configuration yielding the highest silhouette score is identified. Thirdly, for DBSCAN, the algorithm is tuned by adjusting parameters like epsilon and minimum samples, and silhouette scores are calculated for resulting clusters. By evaluating the silhouette scores obtained from each method, the quality of cluster formations is assessed, with higher scores indicative of well-separated clusters. Ultimately, this comparative analysis aids in selecting the most appropriate clustering method and configuration to effectively capture sentiments and preferences expressed in iPhone reviews, facilitating a deeper understanding of customer perceptions regarding various aspects of the product.

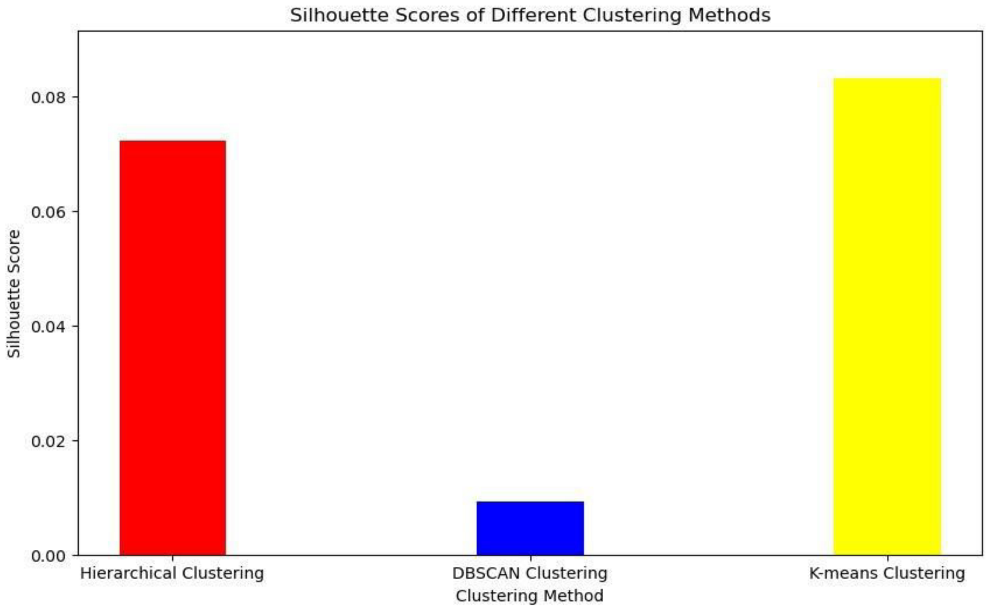

Figure 10: Comparing silhouette scores across different clustering methods

5. CONCLUSION

Aspect-based sentiment analysis (ABSA) presents a nuanced approach to comprehending sentiments expressed in textual data by dissecting it into specific aspects or entities and scrutinizing sentiments associated with each aspect. A comprehensive framework for precise sentiment identification in textual data has been developed, contributing to the advancement of sentiment analysis methodologies by providing a robust solution for understanding nuanced sentiment expressions in textual data. The process of analysing iPhone reviews through clustering techniques yields valuable insights into customer sentiments and preferences concerning the product. Employing methods such as K-means, hierarchical clustering, and DBSCAN enables the effective partitioning of review data into distinct clusters based on similarity in sentiment expressions. Evaluation of clustering performance via metrics such as silhouette scores offers guidance in selecting the most appropriate clustering method and configuration for the dataset. Through this analysis, a comprehensive understanding of prevalent themes and sentiments expressed by customers is attained, facilitating the identification of key aspects of the iPhone product that drive satisfaction or dissatisfaction. Furthermore, visualizations such as histograms of cluster sizes offer additional insights.

REFERENCES

[1]. Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. Effective attention modeling for aspect-level sentiment classification. In Proceedings of COLING.

[2]. Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In Proceedings of ACL.

[3]. Mengting Hu, Shiwan Zhao, H. Guo, Renhong Cheng, and Z. Su. 2019. Learning to detect opinion snippet for aspect-based sentiment analysis.

[4]. Kanwal Ahmed, Muhammad Imran Nadeem et al., Breaking down linguistic complexities: A structured approach to aspectbased sentiment analysis, Journal of King Saud University – Computer and Information Sciences 35 (2023) 101651

[5]. S. Kausar et al., A Sentiment Polarity Categorization Technique for Online Product Reviews, in IEEE Access, vol. 8, pp. 3594-3605, 2020.

[6]. H. Liang, U. Ganeshbabu and T. Thorne, "A Dynamic Bayesian Network Approach for Analysing Topic-Sentiment Evolution," in IEEE Access, vol. 8, pp. 54164-54174, 2020.

[7]. Gagandeep Kaur, Amit Sharma, A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis, Kaur and Sharma Journal of Big Data (2023)

[8]. R. Cobos et al., A Content Analysis System That Supports Sentiment Analysis for Subjectivity and Polarity Detection in Online Courses," in IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 14, no. 4, pp. 177-187, Nov. 2019.

[9]. Li, Y.; Wang, F.; Zhong, S.-h. A More Fine-Grained Aspect– Sentiment–Opinion Triplet Extraction Task. Mathematics 2023, 11, 3165.

[10]. L. Yu et al., Research on Intelligence Computing Models of Fine-Grained Opinion Mining in Online Reviews," in IEEE Access, vol. 7, pp. 116900-116910, 2019.

[11]. Komang Wahyu Trisna et al., Deep Learning Approach for Aspect-Based Sentiment Classification: A Comparative Review, Applied Artificial Intelligence, 36:1, 2014186, 2022

[12]. S. Shayaa et al., "Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges," in IEEE Access, vol. 6, pp. 37807-37827, 2018.

[13]. S. Hu, A. Kumar, F. Al-Turjman, S. Gupta, S. Seth and Shubham, "Reviewer Credibility and Sentiment Analysis Based User Profile Modelling for Online Product Recommendation," in IEEE Access, vol. 8, pp. 26172-26189, 2020.

[14]. S. Zhang and H. Zhong, Mining Users Trust from E-Commerce Reviews Based on Sentiment Similarity Analysis, in IEEE Access, vol. 7, pp. 13523-13535, 2019.

[15]. G. Xu et al., "Sentiment Analysis of Comment Texts Based on BiLSTM," in IEEE Access, vol. 7, pp. 51522-51532, 2019.

[16]. W. Zhao et al., "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 1, pp. 185-197, 1 Jan. 2018.

[17]. Yoon, Y. Jeong and S. Kim, "Detecting a Risk Signal in Stock Investment Through Opinion Mining and Graph-Based Semi-Supervised Learning," in IEEE Access, vol. 8, pp. 161943-161957, 2020.

[18]. Kanwal Ahmed, Muhammad Imran Nadeem, Zhiyun Zheng, Dun Li, Inam Ullah, Muhammad Assam, Yazeed Yasin Ghadi, Heba G. Mohamed, Breaking down linguistic complexities: A structured approach to aspect-based sentiment analysis, Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 8, 101651, 2023, https://doi.org/10.1016/j.jksuci.2023.101651.

[19]. Chifu, Adrian-Gabriel, and Sébastien Fournier. 2023. "Sentiment Difficulty in Aspect-Based Sentiment Analysis" *Mathematics* 11, no. 22: 4647. https://doi.org/10.3390/math11224647

[20]. Kumar, D., Gupta, A., Gupta, V. K., & Gupta, A., Aspect-Based Sentiment Analysis using Machine Learning and Deep Learning Approaches. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 118–138. 2023, https://doi.org/10.17762/ijritcc.v11i5s.6636

[21]. Ganpat Chauhan, Ravi Nahta, Yogesh Meena, Dinesh Gopalani, Aspect based sentiment analysis using deep learning approaches: A survey, Computer Science Review 49(4):100576, DOI: 10.1016/j.cosrev.2023.100576

[22]. Hengyun Li, Bruce X.B. Yu, Gang Li, Huicai Gao, Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews, Tourism Management, Volume 96, 104707, 2023, https://doi.org/10.1016/j.tourman.2022.104707.