Vol-09 Issue 01, January -2025 ISSN: 2456-9348
Impact Factor: 7.936

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

CLARIVIEW

Mrs. Himani Patel
Shruti Katole
Nidhi Satdeve
Nikita Jadhav
Laxmi Sarde
Department Of Information Technology, D.Y. Patil College Of Engineering, Akurdi, Pune

ABSTRACT:
This research presents a machine learning-driven text summarization tool, utilizing Natural Language Processing (NLP) techniques along with the Natural Language Tool-Kit (NLTK) to create a logical and extractive summarization approach. This tool is designed to generate concise, coherent summaries by isolating key points from lengthy input texts or documents. By omitting redundant information, it provides users with a fluent summary that facilitates quick understanding of essential information. Text for summarization can be integrated via web scraping or manually entered, allowing flexibility in data sourcing. Additionally, this research incorporates aspect-oriented sentiment analysis specifically focused on product reviews. By identifying positive and negative product characteristics, this feature serves a dual purpose: it helps consumers make informed decisions while offering valuable insights to manufacturers seeking product improvements. To achieve this, the system builds category specific sentiment and aspect dictionaries, enabling refined and accurate sentiment analysis. This tool's integration of summarization and sentiment analysis provides a streamlined approach for users, offering both concise summaries and in-depth insights into product attributes, thus enhancing its practical applications for varied end-users.

Keywords:
Natural Language Processing, Text summarization, Aspect base Analysis, Web Scraping, Sentiment Analysis, Product Reviews

INTRODUCTION

In the realm of online shopping, users frequently depend on product reviews to make informed decisions. However, analysing numerous reviews manually can become tedious, especially when reviews are extensive or intricate. Summarization techniques provide a solution to this issue by condensing reviews into concise, informative sentences, allowing consumers to quickly assess different product options. This research introduces ReviewSense, a system designed for aspect-based sentiment analysis, leveraging Amazon product reviews to automate the extraction of both sentiments and aspects, which provides meaningful insights for consumers and manufacturers alike. nidhisatdeve@gmail.com

As e-commerce continues to expand rapidly, consumers face an overwhelming array of product choices and information, often leading to indecision. Product reviews online are a powerful resource for informed decision-making but paradoxically contribute to the complexity of choices. This increasing volume of reviews often overwhelms users, creating a need for solutions to streamline decision-making. Advances in artificial intelligence (AI) are addressing this challenge, transforming how users interact with product reviews. This pioneering solution redefines consumer engagement with reviews by employing advanced natural language processing (NLP) techniques to not only reduce lengthy reviews but also determine the overall sentiment and quality of a product. Additionally, it generates compact summaries, enabling consumers to swiftly understand key product aspects, advantages, and drawbacks.

Three main algorithms are fundamental to this system: Random Forest, Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN). Each of these algorithms contributes distinct capabilities to the process. Random Forest is effective in handling complex data and categorizing it, while LSTM and RNN are particularly suited for understanding and processing sequential information—an essential requirement for parsing and summarizing detailed reviews. By integrating AI and these sophisticated algorithms, the system equips consumers to sift through the mass of reviews, facilitating quicker, better-informed decisions in the fast-paced world of e-commerce.

Python libraries like TextBlob and Natural Language Toolkit (NLTK) play a key role in implementing this solution. TextBlob offers a higher-level interface for performing common NLP tasks, while NLTK provides a broad set of tools and resources for more intricate control over a range of NLP tasks, including text summarization.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

BACKGROUND AND RELATED WORK

A.Text Summarization Types

1.Extractive Summarization:

TextRank: This graph-based algorithm selects the most important sentences from a text by ranking sentences based on their interconnectedness. It's widely used for extractive summarization.

LexRank: Similar to TextRank, LexRank uses cosine similarity between sentences and applies PageRank to identify the most relevant sentences.

Latent Semantic Analysis (LSA): LSA analyses relationships between terms and sentences to capture topic relevance and extract key sentences based on semantic similarity. BERT-based Summarization: Pre-trained transformers like BERT (Bidirectional Encoder Representations from Transformers) and other transformers can be fine-tuned to extract important sentences in a text. BERT identifies context-rich sentences that best represent the content.

2.Abstractive Summarization:

Sequence-to-Sequence (Seq2Seq) Models: These models generate summaries by learning input-output mappings. RNNs, GRUs, and LSTMs are commonly used in Seq2Seq models.

Transformers (e.g., BERT, T5, GPT): Transformer models can be fine-tuned for summarization tasks, generating summaries by understanding context and rephrasing content in fewer words.

Pointer-Generator Networks: A combination of extractive and abstractive methods, this network selects relevant content while generating new phrases to produce more fluent and meaningful summaries.

B. Aspect-Based Sentiment Analysis (ABSA) Algorithms

1.Aspect Extraction:

Rule-Based Approaches: Use predefined linguistic rules and lexicons to extract aspects. While straightforward, they may lack flexibility in complex texts.

Dependency Parsing: Identifies relationships between words in a sentence to capture relevant aspects and sentimentbearing words by analyzing syntactic structure.

LSTM and CNN Models: These deep learning models capture word dependencies and sequence patterns to automatically learn aspect terms from data.

Transformers: BERT and other transformer models excel at aspect extraction due to their ability to capture contextual dependencies and nuances.

2.Sentiment Classification:

Support Vector Machines (SVM) and Naïve Bayes: These classic ML algorithms are used for classifying sentiments in simple cases but require feature engineering for best results.

BiLSTM and GRU Models: These deep learning architectures capture word sequences bidirectionally, making them useful for sentiment polarity classification.

Attention Mechanisms: Used to focus on key words associated with sentiment in the context of specific aspects, enhancing the performance of LSTMs or transformers.

Aspect Sentiment Transformers: Fine-tuning BERT or RoBERTa on ABSA tasks allows the model to assign sentiment polarity to specific aspects, improving accuracy in complex reviews.

PROPOSED SYSTEM

In the proposed system, sentiment similarity analysis is implemented using both unsupervised and supervised machine learning algorithms. This approach frames sentiment classification as a problem of learning sentimentspecific word embeddings. Three distinct neural network models have been designed to effectively integrate supervision from text data with sentiment labels. An innovative method is proposed to compute sentiment similarity by bridging the semantic and emotional spaces. The effectiveness of this approach is demonstrated in two NLP tasks: inference of indirect question-answer pairs and prediction of sentiment orientation. Experimental results indicate that sentiment similarity measurement is crucial for achieving satisfactory performance in these tasks. Our approach shows significant improvement over two widely used semantic similarity measures, namely, PMI (Pointwise Mutual Information) and LSA (Latent Semantic Analysis).

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

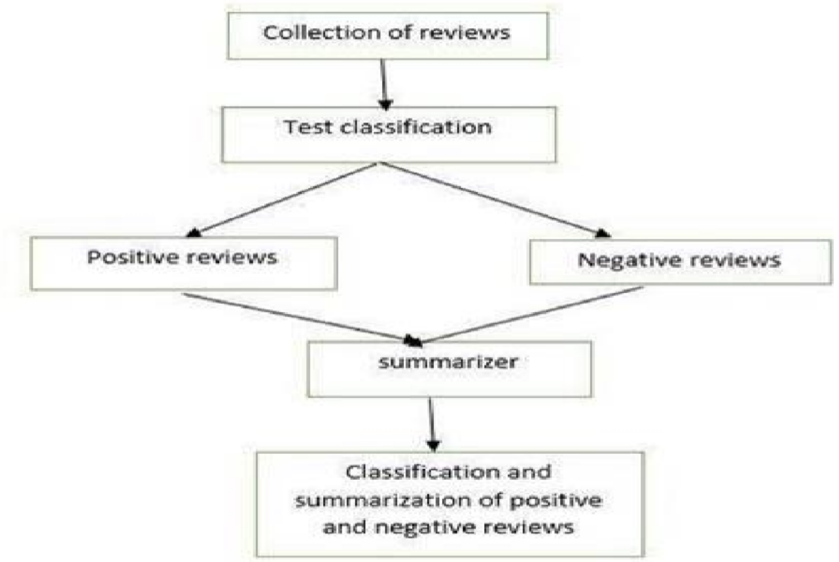

Fig 1 : Block Diagram

Dataset Collection: The dataset used in this system is sourced from Amazon, stored in a .csv file. It includes two main columns: review text and the corresponding rating. With 1,000 entries, this dataset provides ample data for training and analysis.
Data Pre-processing: Pre-processing involves transforming raw data into a format suitable for analysis. Data from various sources are often unstructured and contain noise, making this step essential for cleaning and organizing the data to prepare it for further analysis.
Text Summarization: Text summarization, an area within Natural Language Processing (NLP), focuses on extracting concise summaries from extensive text data. The summarization process includes the following steps: (i) Collecting reviews from Amazon; (ii) Applying preprocessing techniques; (iii) Splitting text into individual sentences; (iv) Calculating weighted word frequency; (v) Scoring sentences; and (vi) Generating a summary.
Training the Data: This step involves training the model on the prepared dataset, allowing it to learn patterns and accurately make predictions. Model performance is assessed using various metrics, which are critical for understanding how well the algorithm has learned from the training data.
Model Prediction and Evaluation: Here, the dataset is divided into a training set for model learning and an independent set for performance evaluation. The evaluation metric, selected based on the specific machine learning task, is classification, providing insights into the accuracy and reliability of the model’s predictions.

LITERATURE SURVEY

[1] B Kavitha Rani1, M Varaprasad Rao2, K Srinivas3, G.Madhukar Professor, CMR Technical Campus, Hyderabad. Text summarization using LSTM deep learning: There are two approaches reported in more recent text summarization research-extractive and abstractive. The latter focuses on key sentence selection - TextRank-whereas the former involves the generation of new sentences to condense the contentabstractive. For low resource languages like Telugu, PENSEE suggested a machine learning-based approach for summarization where the Porter algorithm is used for preprocessing and LSTM networks are used to generate summaries. With regards to accuracy, the approach functioned with a relative standard of 86% where sentence ranking used Cosine Similarity. Challenges in multi-document summarization: Time consumption is a major problem for bigger datasets. It might be solved with deep learning methods by saving time and improving the quality of summaries. Thus, the findings can be applicable in different domains worldwide. The findings will be relevant in improving text summarization tools by these outcomes, particularly for non-English languages.

[2] V. R. Welgamage U. A. C. Senarathne N. H. A. C.
Madhubhashani T. C. Liyanage P. P. G. Dinesh Asanka*
Literature Review: Thorough and Feature-Level Sentiment Analysis of Amazon Product Reviews In this paper, an effort is made towards overall sentiment classification (positive, neutral, negative) as well as feature-level sentiment classification

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

(price, quality, etc.) of a product based on Amazon reviews by using machine learning. Preprocessing includes removing duplicates, tokenizer, removing stopwords and punctuation from the dataset containing 142.8 million reviews including reviewer ID, ratings, and review-text. Sentiment is classified based on ratings (4-5 stars: positive, 3 stars: neutral, 1-2 stars: negative) using a Random Forest model with TF-IDF vectorization, achieving 83.28% accuracy. The architecture includes a Chrome extension for displaying sentiment on Amazon product pages and a Flask server for data processing and real-time model execution. Results are visualized through bar charts and feature-specific insights, combining TF-IDF and VADER for comprehensive sentiment analysis.

[3] Raihah Aminuddin, Aina Zuliana Zulke, Khyrina Airin Fariza Abu Samah: Sentiment analysis of online product reviews using Lexical Semantic Corpus-Based technique

Seq2Seq models are commonly used for text summarization by employing an encoder-decoder architecture, where the encoder processes input sequences (product reviews) and the decoder generates summaries. Using bidirectional LSTMs enhances context understanding by considering both past and future information, which is critical for grasping the nuances in reviews. The Bahdanau attention mechanism helps the model focus on important parts of the review, improving summary relevance. Pre-trained word embeddings like ConceptNet Numberbatch provide semantic understanding, boosting the accuracy of summaries

[4] Dr. R Senthamil Selvi1, Aishwariya P2, Bhagya Lakshmi S3, Darlen Jovita J4, Hamshavarthini M5: Classification and Summarization of Amazon Product Reviews The literature review highlights key techniques used in sentiment analysis and text summarization. Sentiment classification often relies on NLP and lexicon-based approaches to categorize reviews as positive or negative, building on previous research that has achieved over 80% accuracy. Machine learning methods, such as Naive Bayes and Support Vector Machines, have been widely applied to sentiment analysis, with performance depending heavily on pre-processing techniques. For text summarization, strategies like weighted frequency and sentence scoring are employed to generate concise summaries from large volumes of text, such as product reviews.

CHALLENGES AND LIMITATIONS

In Text Summarization and Aspect-Based Sentiment Analysis (ABSA), several challenges and limitations hinder the effectiveness and accuracy of these techniques. Here's a breakdown of some key issues to consider:

Challenges in Text Summarization:

1.Retention of Context and Meaning: Abstractive summarization often struggles to capture nuanced meanings, especially in long or complex texts, resulting in summaries that can lack coherence or miss key points.

2.Handling Redundancy and Irrelevance: Summarization models may include redundant or irrelevant information, particularly in extractive methods, where it's challenging to filter out content that doesn't add value to the summary.

3.Domain-Specific Limitations: Summarization models often generalize poorly across different domains. A model trained on general news articles, for instance, may perform poorly on product reviews, legal documents, or medical text.

4.Computational Resource Intensity: Abstractive models, especially those based on transformers (e.g., BERT, GPT), require substantial computational power, making real-time summarization difficult for large datasets or resourceconstrained environments.

5. Evaluation Metrics: Common metrics like ROUGE focus on word overlap and may not capture the semantic quality or coherence of generated summaries, leading to challenges in accurately evaluating model performance. **Challenges in Aspect-Based Sentiment Analysis:**

1.Aspect Identification and Ambiguity: Extracting relevant aspects from text is complex because the same word may represent different aspects in different contexts (e.g., "battery" in phones vs. laptops), making consistent aspect identification challenging.

2. Contextual Sentiment Classification: The sentiment of an aspect can vary depending on the surrounding context. For instance, the word "lightweight" could be positive for a phone but negative for a camera lens. Capturing such contextual sentiment shifts requires complex modeling.

3. Handling Long and Complex Sentences: Longer sentences with multiple aspects or nuanced expressions pose a challenge for ABSA, as models may overlook sentiment associated with certain aspects or misinterpret relationships within the text.

4.Imbalanced Data: In real-world datasets, certain aspects or sentiment classes may be underrepresented, leading to biased models that perform poorly on less common categories.

5. Implicit Aspect and Sentiment**:** Aspects and sentiments are not always explicitly mentioned, particularly in subjective or conversational language. For instance, "This brand never disappoints" implies positive sentiment but lacks specific aspect terms.

Limitations in Text Summarization and Aspect-Based Sentiment Analysis

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

1.Aspect Identification and Ambiguity: Extracting relevant aspects from text is complex because the same word may represent different aspects in different contexts (e.g., "battery" in phones vs. laptops), making consistent aspect identification challenging.
2.Contextual Sentiment Classification: The sentiment of an aspect can vary depending on the surrounding context. For instance, the word "lightweight" could be positive for a phone but negative for a camera lens. Capturing such contextual sentiment shifts requires complex modeling.
3.Handling Long and Complex Sentences: Longer sentences with multiple aspects or nuanced expressions pose a challenge for ABSA, as models may overlook sentiment associated with certain aspects or misinterpret relationships within the text.
4.Imbalanced Data: In real-world datasets, certain aspects or sentiment classes may be underrepresented, leading to biased models that perform poorly on less common categories.
5.Implicit Aspect and Sentiment: Aspects and sentiments are not always explicitly mentioned, particularly in subjective or conversational language. For instance, "This brand never disappoints" implies positive sentiment but lacks specific aspect terms.

CONCLUSION

In a nutshell, our project on ReviewSense presents a robust solution for automating the product Summarization and Aspect-Based Sentiment Analysis over Effective

Categorization of product reviews into positive, neutral, and negative sentiments, we provide consumers with valuable This implies that such knowledge enables the procurement decisions. Inclusion of feature-level sentiment Now, this method allows businesses to get deep insight into customer feedback over a specific product. attributes include price and quality. Through advanced machine learning techniques and natural language processing methods, ReviewSense transforms unstructured review data into actionable insights, helping businesses identify Areas of improvement and innovation: The summarization tool also summarizes long reviews. into digestible formats, making consumer access to information easier and more responsive. Ultimately, ReviewSense manages to make the difference in an overall customer experience. decision-making process but also provides businesses with essential feedback to enhance product offerings. Looking ahead, we see this system in a prime position to bridge across This can easily link consumer sentiments to business strategies

FUTURE SCOPE

Future work could involve customizing text summarization based on user preferences for summary depth, allowing for more tailored review insights. To enhance accuracy, BERT (Bidirectional Encoder Representations from Transformers) could be integrated for its bi-directional context understanding in NLP tasks. Given BERT's high computational demands, DistilBERT offers a faster alternative, achieving nearly 89% accuracy in our initial tests.
Moreover, topic modelling using LDA (Latent Dirichlet Allocation) could replace predefined features, enabling the system to identify product aspects directly from reviews. Looking forward, this solution could evolve into a recommendation system that highlights related products with highly rated features, aligning recommendations closely with customer preferences.

REFERENCE

[1] B Kavitha Rani1, M Varaprasad Rao2, K Srinivas3, G.Madhukar Professor, CMR Technical Campus, Hyderabad :TEXT SUMMERIZATION USING LSTM DEEP LEARNING
[2] Raihah Aminuddin, Aina Zuliana Zulke, Khyrina Airin Fariza Abu Samah:Sentiment analysis of online product reviews using Lexical Semantic Corpus-Based technique.
[3] V. R. Welgamage U. A. C. Senarathne N. H. A. C. Madhubhashani T. C. Liyanage P. P. G. Dinesh Asanka* :Paper No: SC-31: Overall and Feature Level Sentiment Analysis of Amazon Product Reviews Using Machine Learning Techniques and Web-Based Chrome Plugin
[4] Dr. R Senthamil Selvi1, Aishwariya P2, Bhagya Lakshmi S3, Darlen Jovita J4, Hamshavarthini M5: Classification and Summarization of Amazon Product Review
[5] Sentilyzer: Aspect-Oriented Sentiment Analysis of Product Review:Sentilyzer: Aspect-Oriented Sentiment Analysis of Product Review
[6] Geetha, M.M., & Karthika R. D. "Improving the performance of aspect based sentiment analysis using finetuned BERT base uncased model", 2021.
[7] Y. Xiao, C. Q. & Leng H. "Sentiment analysis of Amazon product reviews based on NLP," 2021, pp.1218-1221, doi: 10.1109/AEMCSE51986.2021.00249.

ISSN: 2456-9348
Impact Factor: 7.936

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

[8] Dr. Gajula Ramesh, Dr.J.Somasekar, Dr. Karanam Madhavi, Dr. Gandikota Ramu, Best keyword set recommendations for building service-based systems International Journal of Scientific and Technology Research, volume 8, issue 10, October, 2019.
[9] Mayuri G. Dabhade and Prof.Nitin Chopde,July 2020 ,A Result Review Analysis of Product Recommendation System in Domain Sensitive Manne
[10] Tripathy, Abinash, Agarwal,A. and Santanu Kumar Rath.: Classification of Sentimental Reviews Using Machine

Learning Techniques. Procedia Computer Science, Vol. 57, pp. 821–829 (2015)