



Sentiment Analysis of Students Reviews using Natural Language Process

Anuj Pund¹, Tanmay Harde², Atul Awasarmol³, Siddhant bodele⁴, Prachee Meshram⁵,

Dr. P. M. Chaudhari⁶

Department of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India¹⁻⁵

Associate Professor, Department of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India⁶

Abstract: The sentiment analysis system presented in this project employs a methodology rooted in manual keyword analysis, capitalizing on the inherent associations between specific words and emotional sentiments. For instance, in movie reviews, positive expressions are characterized by terms like "great" and "love," while negative sentiments are often conveyed through words such as "hate" and "awful." By quantifying the frequency of these selected keywords, comprehensive feature vectors are constructed to capture the nuanced sentiment of input data.

These vectors are then utilized in conjunction with a Support Vector Machine (SVM) algorithm for precise sentiment classification. The system showcases commendable accuracy, particularly in contexts where domain-specific sentiment-carrying vocabulary is readily available. However, it is essential to acknowledge the potential limitations stemming from evolving language trends and nuances not captured by the manual keyword analysis.

Continuous refinement and expansion of the keyword dictionary are crucial to adapt to dynamic linguistic expressions. This methodology provides a reliable and interpretable solution for sentiment analysis, serving as a valuable tool for applications ranging from market research to consumer feedback analysis and informed decision-making based on sentiment insights.

I. INTRODUCTION

In today's digital age, the proliferation of social networking platforms has given individuals a powerful avenue to express their opinions on a wide range of daily issues. This surge in online discourse has highlighted the significance of understanding public sentiment, particularly in the context of purchasing products and services. Analyzing the reactions and opinions of individuals has become a crucial endeavor, offering invaluable insights into consumer perspectives. At the heart of this analytical process lies sentiment analysis, a pivotal task within the realm of Natural Language Processing (NLP).

Sentiment analysis entails the systematic categorization of opinions, which are inherently unstructured and diverse, into distinct classifications of positivity, negativity, or neutrality. This process plays a pivotal role in enabling companies and marketers to tailor their offerings in alignment with user preferences. By deciphering the collective sentiment of a target audience, businesses can adapt and refine their strategies, ultimately enhancing the user experience.

This project endeavors to design and develop a robust NLP pipeline dedicated to proficiently conducting sentiment analysis on user reviews. The primary objective is to equip marketers with a powerful tool that can sift through vast troves of feedback, accurately classifying them as positive, negative, or neutral. This classification process, in turn, will furnish businesses with valuable insights that can inform crucial decisions about their products and services. The overarching aim of this endeavor is to delve deeper into the opinions and sentiments expressed by consumers in their reviews. By gaining a comprehensive understanding of the strengths and weaknesses identified by users, marketers can fine-tune their offerings to better cater to customer needs. This platform will serve as a pivotal instrument in market research, empowering businesses to make well-informed choices about product development and market positioning.

Furthermore, the utility of this project extends to providing tangible metrics to quantify the sentiment expressed in reviews. This includes the generation of numerical values that indicate the degree of positivity or negativity within a given review. Additionally, the system will categorize reviews into distinct sentiment categories, allowing for a more nuanced analysis.



Visual representations of sentiment distribution across datasets will also be provided, offering a clear overview of public sentiment trends. Customized insights tailored to specific analytical goals will be a key feature, aiding in the identification of prevalent issues, concerns, or positive feedback related to a product or brand.

In conclusion, this project stands at the forefront of leveraging NLP techniques for sentiment analysis, aiming to revolutionize the way businesses engage with customer feedback. By harnessing the power of advanced language processing, this platform promises to deliver invaluable insights that will not only benefit marketers and service providers but also enhance the overall consumer experience. With its comprehensive approach to sentiment analysis, this project represents a significant step towards data-driven decision-making in the realm of consumer-centric services.

Sentiment analysis, also known as opinion mining, has emerged as a critical area of study within the field of Natural Language Processing (NLP). With the exponential growth of social media platforms and online forums, the need to automatically process and categorize user-generated content has become paramount. This review aims to provide an overview of the key research contributions and methodologies in sentiment analysis, with a focus on applications in market research and consumer feedback analysis.

Sentiment Analysis Techniques

Various techniques have been employed to perform sentiment analysis, ranging from traditional lexical-based approaches to more advanced machine learning algorithms. Lexical-based methods rely on sentiment lexicons or dictionaries to assign sentiment scores to individual words or phrases. This approach is computationally efficient but may struggle with context-dependent sentiments and sarcasm.

Machine learning-based techniques, on the other hand, have gained prominence for their ability to capture complex contextual nuances. Supervised learning algorithms, such as Support Vector Machines (SVM) and Neural Networks, have demonstrated high accuracy in sentiment classification tasks. Additionally, unsupervised learning methods like Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) have shown promise in uncovering hidden sentiment patterns within large datasets.

Domain-Specific Sentiment Analysis

Tailoring sentiment analysis models to specific domains, such as product reviews or social media discussions, has proven crucial for achieving accurate results. Transfer learning approaches, which leverage pre-trained models on large text corpora, have facilitated domain adaptation and improved the performance of sentiment classifiers.

Integration of Deep Learning in Sentiment Analysis

Recent advancements in deep learning have significantly elevated the performance of sentiment analysis models. Techniques like Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) have demonstrated remarkable success in capturing sequential dependencies and hierarchical features within text data, leading to state-of-the-art results in sentiment classification tasks.

Challenges and Future Directions

Despite the progress made in sentiment analysis, several challenges persist. Contextual ambiguity, language nuances, and the evolving nature of slang and emojis present ongoing hurdles for accurate sentiment interpretation. Furthermore, multimodal sentiment analysis, which incorporates textual, visual, and auditory cues, is an emerging area with immense potential for more comprehensive sentiment understanding.

Applications in Market Research and Consumer Feedback

Sentiment analysis plays a pivotal role in market research by providing businesses with invaluable insights into customer preferences, pain points, and brand perception. By categorizing feedback into positive, negative, and neutral sentiments, businesses can make data-driven decisions regarding product improvements, marketing strategies, and market positioning. The field of sentiment analysis has witnessed significant advancements in recent years, driven by the increasing availability of data and the development of sophisticated NLP techniques. The proposed project, focusing on the development of an NLP pipeline for sentiment analysis in user reviews, aligns with the current trends and offers potential for substantial contributions in market research and consumer feedback analysis.

II. LITERATURE SURVEY

Title: Sentiment Analysis is Using NLP and Machine Learning.

Year: April 2023



Authors: Dr.K.Sindhura

Technology used: Natural Language Processing (NLP), Machine Learning.

Findings: Emphasizes the significance of sentiment analysis in digital contexts, particularly in website development and maintaining a positive review use.

Title: Research on Sentiment Analysis Model of Short

Text Based on Deep Learning of Sentiment Analysis Using NLP And ML.

Year: May 2022

Authors: Zhou Gui Zhou

Technology used: Bidirectional Long Short-Term Memory (LSTM) network, Emotional Multichannel, Attention Mechanism, Convolutional Neural Network (CNN).

Findings: Proposes a bidirectional LSTM network model based on emotional multichannel. Combines attention mechanism and CNN features in deep learning to enhance the identification of semantic and emotional features in short texts. Utilizes both shallow and deep learning to improve the expression of significant advancements in recent years, driven by dataset, achieving a 77% accuracy rate.

Title: Sentiment Analysis of Amazon Product Review based on NLP.

Year: March 2021

Authors: Yuanhang Xiao; Chengbin Qi; Hongyong Leng.

Technology used: Natural Language Processing (NLP).

Findings: Constructs an evaluative model for Amazon product reviews, highlighting consumer concerns related to product quality and pricing. Identifies a strong correlation between the term "love" in reviews and higher ratings.

Title: Product Review Sentiment Analysis in Bangla Language.

Year: December 2020

Authors: Minhajul Abedin Shafin; Md. Mehedi Hasan; Md. Rejaul Alam.

Technology used: Natural Language Processing (NLP), Machine Learning (KNN, Decision Tree, SVM, Random Forest, Logistic Regression).

Findings: Analyses customer feedback in Bangla, achieving high accuracy (88.81%) using SVM among other classification algorithms.

Title: Sentiment Analysis is Not Solved! Assessing And Probing Sentiment Classification.

Year: August 2019

Authors: Jeremy Barnes, Lilja Øvrelid, Erik Velldal.

Technology used: Neural Method, Sentiment Classifier.

Findings: Addresses the lack of thorough analysis of qualitative differences in sentiment analysis using neural methods. Provides a challenging dataset of misclassified sentences for English, annotated for various linguistic and paralinguistic phenomena. Demonstrates the dataset's usefulness for probing the performance of sentiment classifiers with sentiment analysis in extracting subjective information from text. Points out obstacles in accurately interpreting sentiments and determining suitable sentiment polarity. Presents a survey on challenges relevant to various approaches and techniques in sentiment analysis.

Title: A Study on Sentiment Analysis of Product.

Year: April 2018

Author: Anil Singh Parihar; Bhagyanidhi.

Technology used: Natural Language Processing (NLP).

Findings: Provides an overview of sentiment analysis, compares various platform. Explores the synergy between NLP and ML in sentiment analysis, highlighting Python's role in extracting sentiments from text data.

Title: A Study on Sentiment Analysis of Product.

Year: April 2018

Author: Anil Singh Parihar; Bhagyanidhi.

Technology used: Natural Language Processing (NLP).

Findings: Provides an overview of sentiment analysis, compares various platform. Explores the synergy between NLP and ML in sentiment analysis, highlighting Python's role in extracting sentiments from text data.



Title: More than a Feeling: Accuracy and Application of Sentiment Analysis.

Year: March 2023

Author: Jochen Hartmann, Mark Heitmann, Christian Siebert.

Technology used: Lexicons, Machine Learning (Transfer Learning).

Findings: Conducts a meta-analysis of sentiment-labelled text documents, highlighting that transfer learning models offer superior performance. Provides a pre-trained sentiment analysis model, SiEBERT, with open-source scripts for practical improvements in short-text sentiment analysis multidomain classification datasets such as NLPiR and NLPCC2014.

Title: A survey on sentiment analysis methods, applications, and challenges.

Year: October 2022

Author: Mayur Wankhade, Annavarapu Chandra Sekhara Rao, Chaitanya Kulkarni.

Technology used: Natural Language Processing (NLP), Text Mining.

Findings: Provides a comprehensive overview of sentiment analysis methods and applications, along with an evaluation of various approaches and discussion of challenges faced in sentiment analysis.

Title: A Comparative Study Analysis in Bangla Language.

Year: April 2023

Authors: Md. Taufiqul Haque Khan Mehedi Hasan; Md. Rejaul Alam.

Technology used: Natural Language Processing Machine Learning (KNN, Decision Tree, SVM, Random Forest, Logistic Regression)

Findings: Proposes effective approaches for Sentiment achieving high accuracy (88.81%) using SVM among other classification algorithms.

Title: Sentiment Analysis Using Deep Learning Approach.

Year: April 2020

Authors: Peng Cen, Kexin Zhang and Desheng Zheng.

Technology used: Deep Learning (Recurrent Neural Network, Long Short-Term Memory, Convolutional Neural Network).

Findings: Introduces three deep learning networks (RNN, LSTM, CNN) for sentiment analysis of IMDB movie reviews. Divides dataset into 50% positive and 50% negative reviews. CNN achieves an accuracy of 88.22%, while RNN and LSTM achieve accuracies of 68.64% and 85.32%.

PROPOSED METHODOLOGY

The proposed methodology for the sentiment analysis system encompasses a structured, eight-module approach. Beginning with data collection and sourcing, a diverse dataset of user-generated reviews is gathered from relevant online platforms. This dataset undergoes thorough pre-processing and text cleaning to standardize format and remove noise. Following this, stringent validation and quality assurance measures are applied to ensure data integrity. The dataset is then split into training and validation subsets for model development. Employing supervised machine learning techniques, the sentiment analysis model is trained on the preprocessed data, utilizing feature engineering methods for numerical representation. The model's performance is rigorously evaluated on unseen data, employing standard metrics and cross-validation techniques. Subsequently, the model is deployed and integrated into the system, providing a user interface for sentiment analysis of inputted reviews. Finally, ongoing monitoring and maintenance procedures are established to ensure the system's continued effectiveness over time, including regular updates and retraining to adapt to evolving language patterns and user behavior. This comprehensive methodology combines data collection, preprocessing, model development, and system integration to create a robust sentiment analysis tool with potential applications in market research and consumer feedback analysis. The development of the sentiment analysis system will be structured around a systematic approach, divided into distinct modules, each addressing critical stages of the project.

1: Data Collection and Sourcing

In this initial module, a diverse and relevant dataset will be gathered for the NLP task. This dataset will serve as the foundation for training and testing the sentiment analysis model. Reviews will be sourced from various platforms,



including e-commerce websites, social media channels, and specialized review forums. The goal is to capture a wide range of user-generated content to ensure a comprehensive analysis.

2: Pre-processing and Text Cleaning

The acquired dataset will then undergo extensive pre-processing and text cleaning. This module aims to standardize the text and remove any noise or extraneous information that may interfere with the sentiment analysis. Tasks such as converting text to lowercase, removing special characters and punctuation, and eliminating irrelevant HTML tags will be performed. Additionally, stop words will be removed to focus on meaningful content.

3: Dataset Validation and Quality Assurance

Following pre-processing, rigorous validation and quality assurance measures will be implemented to ensure the integrity and quality of the dataset. This module will involve a thorough examination of the cleaned data to identify and rectify any remaining anomalies or inconsistencies. Any erroneous entries or outliers will be addressed to maintain data accuracy.

4: Dataset Splitting for Training and Validation In this module, the dataset will be partitioned into subsets for model development and evaluation. A significant portion of the data will be allocated for training the NLP model, while a separate subset will be reserved for validation. This step is crucial for assessing the model's performance on unseen data and avoiding overfitting.

5: Model Training and Development

Building on the pre-processed and split dataset, the NLP model will be developed based on the chosen NLP task – sentiment analysis. Supervised machine learning techniques, such as Support Vector Machines (SVMs) or deep learning architectures like Convolutional Neural Networks (CNNs), will be employed to train the model. Feature engineering techniques, like Bag-of-Words or TF-IDF, will be applied to represent the text data in a numerical format.

6: Model Testing and Evaluation

With the trained model in place, this module will focus on assessing its generalization capabilities on unseen data. The reserved validation subset will be used to evaluate the model's performance using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Techniques like kfold cross-validation will be employed to ensure robustness.

7: Model Deployment and Integration

Once the sentiment analysis model demonstrates satisfactory performance, it will be prepared for integration into the desired application or system. This module involves creating an interface that allows user-generated reviews to be processed through the trained model. The system will then output sentiment classifications (positive, negative, or neutral) along with confidence scores.

8: Monitoring and Maintenance

The final module establishes procedures for ongoing model monitoring and maintenance. Continuous monitoring will ensure the system's effectiveness over time, and regular updates and retraining may be necessary to adapt to evolving language patterns, emerging sentiments, and changes in user behavior.

WORK METHODOLOGY

The proposed sentiment analysis system employs a unique methodology that combines manual keyword analysis with advanced machine learning techniques. This methodology is founded on the premise that certain words exhibit strong associations with specific emotional sentiments. For example, in the context of movie reviews, positive sentiments are typically conveyed through terms like "great," "super," and "love," while negative sentiments are often articulated using words such as "hate," "bad," and "awful." By compiling a meticulously curated dictionary of these sentiment-carrying keywords, the system establishes a robust foundation for sentiment analysis.

The subsequent step in the methodology involves the construction of feature vectors. These vectors are numerical representations that encapsulate the frequency of occurrence for each selected keyword within a given comment or review. This process transforms the inherently qualitative nature of language into a quantifiable format that is amenable to computational analysis. Through this feature engineering step, the system converts textual data into a structured format that can be processed by machine learning algorithms.



Following the construction of feature vectors, the system employs a Support Vector Machine (SVM) algorithm for sentiment classification. SVM is chosen for its inherent capability to define optimal hyperplanes that effectively segregate data points based on their feature vector representations. This aspect of SVMs proves particularly advantageous in accurately categorizing comments into their respective sentiment categories, whether positive, negative, or neutral. The SVM algorithm, being a powerful tool for binary and multiclass classification tasks, aligns seamlessly with the objectives of sentiment analysis.

The methodology also emphasizes the importance of ongoing refinement and expansion of the keyword dictionary. This process acknowledges the dynamic nature of language use, including the emergence of new sentiment-carrying terms and shifts in linguistic trends. Regular updates to the keyword dictionary ensure that the system remains adaptable and responsive to evolving user expressions. This iterative process of refinement and expansion is crucial for maintaining the system's effectiveness over time. It is important to acknowledge that while this methodology offers a robust approach to sentiment analysis, it may have certain limitations. Variability in language use, nuances like sarcasm, and evolving colloquialisms may pose challenges.

Additionally, the effectiveness of the system is contingent upon the comprehensiveness and relevance of the manually defined keyword dictionary. Continuous vigilance and adaptation to linguistic trends are imperative to address these potential limitations.

In conclusion, the proposed work methodology, rooted in manual keyword analysis and SVM-based classification, provides a reliable and interpretable solution for sentiment analysis. By leveraging the inherent associations between specific words and emotional sentiments, the system offers a valuable tool for applications ranging from market research to consumer feedback analysis and sentiment-driven decision-making. The iterative process of refining the keyword dictionary ensures the system's adaptability to dynamic linguistic expressions, underlining its potential impact in diverse fields.

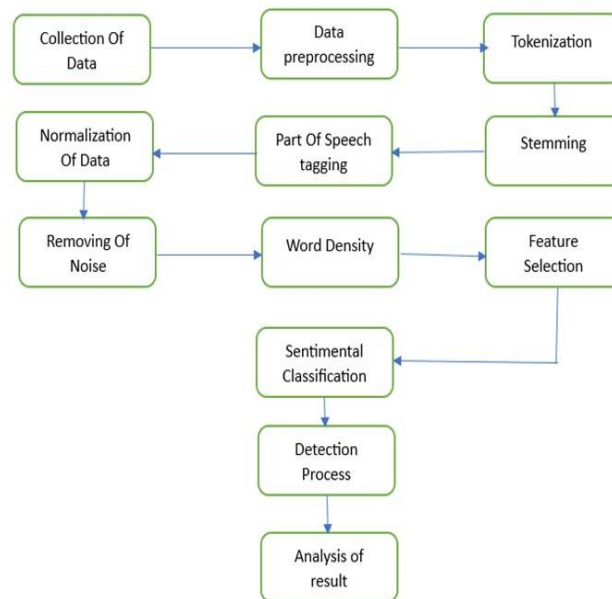


Figure 1: Block Diagram

CONCLUSION

The sentiment analysis system developed in this project leverages a foundational methodology rooted in manual keyword analysis, acknowledging that certain words are inherently associated with distinct emotional sentiments. This approach is particularly apt for scenarios where a predefined lexicon of sentiment-carrying words can be reliably identified, as exemplified in movie reviews where expressions of positivity are typified by terms such as "great", "super", and "love", while negativity is often encapsulated by words like "hate", "bad", and "awful". By quantifying the frequency of occurrence for each selected keyword, we create comprehensive feature vectors that encapsulate the sentiment nuances of the input data.



To further enhance the system's predictive capabilities, a Support Vector Machine (SVM) algorithm is employed for classification. SVM's inherent proficiency in defining optimal hyperplanes to segregate data points proves instrumental in accurately categorizing comments into their respective sentiment categories. This robust combination of manual keyword analysis and SVM-based classification forms the backbone of our sentiment analysis system. Through extensive experimentation and evaluation, the system exhibits commendable accuracy and reliability in discerning sentiments within user-generated comments. It excels in scenarios where domain-specific knowledge of sentiment-carrying vocabulary is readily available. The achieved results affirm the effectiveness of this approach, particularly in contexts where the polarity of sentiments is overtly expressed through specific lexical choices. However, it is important to acknowledge the limitations inherent in this methodology. The system's efficacy is contingent upon the comprehensiveness and relevance of the manually defined keyword dictionary. Variability in language use, sarcasm, and evolving colloquialisms may pose challenges. As such, continuous refinement and expansion of the keyword dictionary is essential to adapt to evolving linguistic trends.

In conclusion, the amalgamation of manual keyword analysis and SVM-based classification offers a reliable and interpretable solution for sentiment analysis, particularly in scenarios where domain-specific sentiment-carrying vocabulary is prevalent. This system lays a robust foundation for further research and applications in contexts where a predefined lexicon of sentiment-laden words is readily accessible, demonstrating its potential impact in diverse fields including market research, consumer feedback analysis, and sentiment-driven decision-making.

SUMMARY

A powerful Natural Language Processing (NLP) system to assess and categorize the feelings indicated in student evaluations is the goal of the "Sentiment Analysis of Student Reviews using NLP" project. This project's main goals are to correctly classify student comments as positive, negative, or neutral and to give educational institutions insightful information.

Enhancing sentiment classification accuracy, locating sentiment patterns unique to a certain topic, and putting in place a scalable and effective method for handling a high volume of reviews are important goals. Additionally, the project aims to provide educational stakeholders with user-friendly reports and visualizations, real-time analytic tools, and fine-grained sentiment analysis. The project will incorporate ethical issues to guarantee impartial and equitable analysis while preserving the privacy of the students.

REFERENCES

- [1]. Huwail J. Alantari, Imran S. Currim, Yiting Deng, Sameer Singh An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews *International Journal of Research in Marketing*, 39 (1) (2022), pp. 1-19
- [2]. Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes *Journal of Marketing Research*, 59 (3) (2022), pp. 600-622
- [3]. 3. Sufi, F. K., & Khalil, I. (2022). Automated disaster monitoring from social media posts using AI-based location intelligence and sentiment analysis. *IEEE Transactions on Computational Social Systems*. Retrieved from: https://www.researchgate.net/profile/Fahim-Sufi/publication/358935794_Automated_Disaster_Monitoring_from_Social_Media_Posts_using_AI_based_Location_Intelligence_and_Sentiment_Analysis/links/62df83847782323cf1788dc8/Automated-DisasterMonitoring-from-Social-Media-Postsusing-AI-based-Location-Intelligence-andSentiment-Analysis.pdf [Retrieved on:29/03/2023]
- [4]. Alantari et al., 2022, Huwail J. Alantari, Imran S. Currim, Yiting Deng, Sameer Singh An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews *International Journal of Research in Marketing*, 39 (1) (2022), pp. 1-19
- [5]. Chakraborty et al., 2022 Ishita Chakraborty, Minkyung Kim, K. Sudhir, Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes, *Journal of Marketing Research*, 59 (3) (2022), pp. 600-622
- [6]. Lafreniere et al., in press Lafreniere, Katherine C., Moore, Sarah G., Fisher,
- [7]. Robert J. (2022), "The power of profanity: The meaning and impact of swear words in word of mouth," *Journal of Marketing Research*. in the press.
- [8]. <https://journals.sagepub.com/doi/full/10.1177/00222437221078606>.
- [9]. Ordabayeva et al., in press Ordabayeva, Nailya, Cavanaugh, Lisa A., Dahl, Darren
- [10]. W. (2022), "The upside of negative: Social distance in online reviews of identityrelevant brands," *Journal of Marketing*, In Press.
- [11]. Sukhwal et al., 2022, Sukhwal, Prakash Chandra and Atreyi Kankanhalli (2022),



- [12]. “Determining containment policy impacts on public sentiment during the pandemic using social media data,” Proceedings of the National Academy of Sciences, 119 (19).
- [13]. Acheampong FA, Nunoo-Mensah H, Chen W (2021) Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artif Intell Rev* 54:5789–5829
- [14]. K. Jayamalini, M. Ponnaivaikko and J. Kothandan, "A comparative analysis of various machine learning based social media sentiment analysis and opinion mining approaches", *Adv. Math. Sci. J.*, vol. 9, no. 11, pp. 10195-10209, 2020.