Republic of the Philippines
CAVITE STATE UNIVERSITY
Don Severino de las Alas Campus
Indang, Cavite

SENTIMENT ANALYSIS OF FACEBOOK POST COMMENTS USING NAIVE BAYES CLASSIFIER AND SEMANTIC ROLE LABELING

Rationale/ Introduction

Sentiment analysis involves identifying and extracting the emotions and attitudes expressed within text. This method allows for the assessment of whether a sentence conveys a positive, negative, or neutral sentiment. Additionally, the integration of Naive Bayes Classifier into sentiment analysis systems has been well-established in existing studies and literature. In fact, studies like Bayhaqy et al. (2018) have demonstrated that the Naive Bayes Classifier is the best classifier for use with social media datasets because it provides more accurate and precise predictions. Naive Bayes Classifier is a widely adopted machine learning algorithm known for its simplicity, efficiency, and effectiveness in categorizing text data into predefined sentiment categories. However, despite the widespread use in text data categorization due to its simplicity and efficiency, it lacks semantic understanding that leads to challenges in accurately interpreting sentiments from longer texts and handling contextual ambiguity (Tesoro et al., 2020). This gap not only obstructs the effectiveness of sentiment analysis systems but also inhibits the extraction of meaningful insights from vast amounts of user-generated content.

To mitigate the aforementioned research gap, the proposed study aims to integrate the strength of Naive Bayes Classifier and Semantic Role Labeling (SRL) techniques into sentiment analysis systems. Semantic role labeling or shallow semantic parsing assigns labels to words or phrases in sentences to identify "Who," "What," "Whom," "How," "When," and "Where" actions occur. Its aim is to recognize events, participants, and properties within sentences by determining the semantic roles of predicates (Shi & Lin, 2019). By combining these approaches, the sentiment analysis system can not only discern the polarity of sentiments but also identify the semantic roles and relationships within text. This provides more nuanced and contextually relevant sentiment predictions. This integration of complementary methodologies offers a promising avenue for improving the accuracy and depth of sentiment analysis in Facebook post comments, enhancing the ability to extract valuable insights from online discourse.

Along with this, the development of more sophisticated sentiment analysis techniques can contribute to advancements in natural language processing and computational linguistics. Moreover, in practical applications such as market research, political analysis, and public opinion monitoring, the proposed approach can offer insights into consumer behavior, societal

trends, and public sentiments on various issues. By providing more accurate and insightful sentiment analysis, this research has the potential to inform decision-making processes and facilitate more informed interventions in both online and offline domains.

All in all, this study lies in the potential of combining the strengths of Naive Bayes Classifier and Semantic Role Labeling to enhance the accuracy of sentiment analysis on social media platforms like Facebook. By integrating these techniques, the sentiment analysis system can capture not only the sentiment polarity but also the semantic roles and relationships within text to provide more insightful and contextually relevant sentiment predictions in the digital age.

Significance of the Study

The proposed study holds significant importance and relevance in today's digital age. As social media platforms like Facebook continues to grow, they serve as a vast repositories of user-generated content including opinions, ideas, and sentiments form a diverse range of individuals. By conducting sentiment analysis of Facebook post comments, this study contributes to unlocking valuable insights into public sentiment dynamics and societal trends.

Insights derived from sentiment analysis can inform evidence-based decision-making processes, policy formulation, and strategic planning in areas such as public health, education, and marketing. Moreover, the integration of machine learning techniques such as Naive Bayes classifier and semantic role labeling (SRL) into sentiment analysis systems represents a methodological advancement with broader implications. These techniques offer not only enhanced accuracy in sentiment prediction but also deeper insights into the semantic structure of language. The findings of this research can serve as a valuable resource for other researchers in the field of natural language processing and machine learning.

In terms of its contribution to the United Nations Sustainable Development Goals (SDGs), this study aligns with SDG 9: Industry, Innovation, and Infrastructure. By advancing sentiment analysis techniques, the research promotes technological progress and nurture innovation. Through the development and application of machine learning techniques for sentiment

analysis, the study contributes to building resilient infrastructure and promoting sustainable industrial development, thus addressing the objectives of SDG 9. Additionally within the context of CvSU Research Thematic Area involvement, the study's focus on Smart Engineering, ICT, and Industrial Competitiveness shows its commitment to enhancing technological capabilities and driving innovation within the academic community.

Scope and Limitations of the Study

This study focuses on enhancing the accuracy and depth of sentiment analysis on Facebook post comments by integrating the Naïve Bayes Classifier and Semantic Role Labeling (SRL) techniques. It involves the development of a sentiment analysis system capable of classifying comments into positive, negative, or neutral sentiments. The system will be trained and validated using a dataset of Facebook post comments to ensure its effectiveness. Additionally, the study evaluates the system’s performance through key metrics such as accuracy, precision, recall, and F1-score while analyzing the impact of integrating Naïve Bayes and SRL on improving sentiment classification.

However, the study has several limitations that must be considered. First, the accuracy of the sentiment analysis system is dependent on the quality and diversity of the training dataset; biases in data collection may affect its generalizability to different contexts. Second, while the Naïve Bayes Classifier and SRL enhance sentiment classification, they may not fully capture complex linguistic nuances such as sarcasm, idiomatic expressions, or evolving online slang, which can lead to misclassification. Additionally, the study is limited to Facebook post comments, meaning that the findings may not be directly applicable to other social media platforms with different user behaviors and language patterns.

Despite these limitations, this research contributes to the advancement of sentiment analysis by demonstrating the effectiveness of integrating machine learning and semantic analysis techniques. The findings provide valuable insights into improving text classification models, which can be utilized in social media monitoring, brand analysis, and opinion mining. Future studies may address the limitations by incorporating deep learning models, expanding the dataset across multiple platforms, and exploring additional linguistic features for more comprehensive sentiment analysis.

Objectives of the Study

The general objective of the study is to enhance the accuracy and depth of sentiment analysis on Facebook post comments by integrating Naive Bayes Classifier and Semantic Role Labeling (SRL) techniques. Specifically, the study aims to:

1. To develop an enhanced sentiment analysis system capable of accurately categorizing Facebook post comments into positive, negative, or neutral sentiments.
2. To train and validate the sentiment analysis system using a dataset of Facebook post comments to ensure its effectiveness.
3. To assess the performance metrics of the sentiment analysis system including accuracy, precision, recall, and F1-score.
4. To analyze the effectiveness of integrating Naive Bayes Classifier and Semantic Role Labeling (SRL) techniques in improving sentiment analysis accuracy and capturing semantic structures within text.

Expected Outputs

The system begins by preparing the text data for analysis. This involves cleaning the text by removing unnecessary characters, punctuation, and symbols to ensure clarity and coherence. Then, the text will undergo tokenization, where it is broken down into individual words or tokens for further examination. Normalization follows by standardizing the text by converting it to lowercase to maintain consistency in analysis regardless of capitalization variations. Stop words such as common and insignificant words like "the" or "and," are then removed to focus on meaningful content. Lastly, lemmatization or stemming is applied to reduce words to their root forms. Once the text is preprocessed, the system will employ a Naive Bayes Classifier to determine the sentiment of the comments. This classifier learns from

a labeled dataset of Facebook comments to understand the patterns associated with positive, negative, and neutral sentiments. By extracting relevant features from the training data, the classifier becomes proficient in predicting the sentiment of new comments based on the learned patterns.

Furthermore, the system integrates Semantic Role Labeling (SRL) to gain insights into the contextual relationships within the comments. Through dependency parsing, it analyzes the grammatical structure of sentences and identifies relationships between words, such as subject-verb-object relationships. Semantic roles including entities such as agents, patients, times, or locations are then assigned to words or phrases within the comments. By integrating this contextual information with sentiment analysis, the system will achieve a deeper understanding of the emotions expressed in the comments.

To implement the system effectively, several materials are required including a dataset of labeled Facebook post comments, Python libraries for natural language processing tasks, a machine learning library like scikit-learn, and an SRL toolkit such as Stanford CoreNLP or spaCy.

References

Concepcion, A., & Sy, C. (n.d.). Modeling the spread of fake news on social networking sites using the system dynamics approach. ASEAN Engineering Journal. https://journals.utm.my/aej/article/view/19251

Duzen, Z., Riveni, M., & Aktaş, M. S. (2023). Analyzing the spread of Misinformation on Social Networks: A process and software architecture for detection and analysis. Computers, 12(11), 232. https://doi.org/10.3390/computers12110232

Luo, H., Cai, M., & Cui, Y. (2021, December 16). Spread of misinformation in social networks: Analysis based on Weibo Tweets. Security and Communication Networks. https://www.hindawi.com/journals/scn/2021/7999760/

Simulation of social media networks with agent-based modeling (2019). https://www.theseus.fi/bitstream/handle/10024/174027/Frida-Gullichsen-thesis.pdf?sequence=2&isAllowed=y

Sulis, E., & Tambuscio, M. (2020). Simulation of misinformation spreading processes in social networks: an application with NetLogo. https://ieeexplore.ieee.org/document/9260064

T, S. M., & Mathew, S. K. (2022, February 15). The disaster of misinformation: A Review of Research in Social Media - International Journal of Data Science and Analytics. SpringerLink. https://link.springer.com/article/10.1007/s41060-022-00311