ABSTRACT

This paper introduces an integrated framework to analyze WhatsApp chat data by coupling sentiment analysis with computational social science techniques. Using social set analysis, which combines both text and network analysis, it extracts valuable information from datasets comprising over 10,000 messages across multiple chat groups. Implemented using lightweight Python libraries (such as NLTK v3.8 and TextBlob v0.17.1) and a Flutter-based user interface, the tool provides an efficient method for users to visualize sentiments and explore chat data. The system achieves an average sentiment classification accuracy of 88.5%, with real-time processing capabilities that handle up to 500 messages per second. The key features include sentiment classification using Python libraries, ensuring reliable results with an F1-score of 0.86 across benchmarked datasets. The Flutter interface improves usability by reducing user interaction time by approximately 30% compared to traditional command-line methods, enabling non-technical users to interact seamlessly with the data. Network analysis modules enable the visualization of communication networks, revealing key metrics such as degree centrality and clustering coefficients with datasets involving up to 100 participants per group. Case studies demonstrate the tool's versatility in detecting sentiment shifts with 93% accuracy during specific events and in identifying emotional responses to contextual triggers, such as public announcements or crises. This framework provides a dynamic platform for researchers and social scientists to interpret conversations, analyze group behavior, and understand emotional exchanges in digital communication platforms. Additionally, it lays the groundwork for integrating advanced machine learning models, such as LSTM and BERT, contributing openings for further investigation in NLP, communal computing, and sentiment analysis.

List of Acronyms

BER	Bit Error Rate
CDMA	Code Division Multiple Access
VADER	Valence Aware Dictionary and sentiment Reasoner
NLTK	Natural Language Toolkit
BERT	Bidirectional Encoder Representations from Transformers

List of Symbols

K	Accuracy score
p1	Proportion of correctly classified messages
pe	Expected proportion of correct classifications (baseline accuracy)
N	Total number of messages (or instances)
y	True label (0 or 1)
y1	Predicted probability of the positive class
TP	True Positives (correctly predicted positive cases)

List of Figures

Fig. 2.1	Sentimental Architecture	13
Fig .3.1	Flow of Process	17
Fig .4.1	Algorithms based on accurancy	30
Fig .4.2	Comparison of Algorithms based on Efficiency, scalability and Accuracy	31
Fig .4.3.	Sentimental Analysis (Txt_1)	32
Fig 4.4	Emoji Analysis (Txt_1)	34
Fig 4.5	Number of occurred words (Txt_1)	35
Fig.4.6	Top Statistics(Txt_1)	36
Fig .4.7.	Sentimental Analysis (Txt_2)	37
Fig 4.8	Emoji Analysis (Txt_2)	38
Fig.4.9	Top Statistics(Txt_2)	39
Fig 4.10	Number of occurred words (Txt_2)	40

List of Tables

1.1	Literature Review and Related work	2-3
2.4	Words Analysis	15