OPINION MINING AND SENTIMENT PROPAGATION

Project Submitted to the
SRM University AP, Andhra Pradesh
for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology
in
Computer Science & Engineering
School of Engineering & Sciences

submitted by

Naga Charitavya Madala(AP21110010027)

Venkata Srikari Malladi(AP21110010036)
Sai Durga Saradhi Pranu Deepak Tallapudi(AP21110011173)

Under the Guidance of

Prof. Rajiv Senapati

Department of Computer Science & Engineering
SRM University-AP
Neerukonda, Mangalgiri, Guntur
Andhra Pradesh - 522 240
May 2025

DECLARATION

I undersigned hereby declare that the project report **Opinion Mining and Sentiment Propagation** submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology in the Computer Science & Engineering, SRM University-AP, is a bonafide work done by me under supervision of Prof. Rajiv Senapati. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree of any other University.

Place	:	____________	Date	:	April 25, 2025
Name of student	:	Naga Charitavya Madala	Signature	:	____________
Name of student	:	Venkata Srikari Malladi	Signature	:	____________
Name of student	:	Sai Durga Saradhi Pranu Deepak Tallapudi	Signature	:	____________

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SRM University-AP

Neerukonda, Mangalgiri, Guntur

Andhra Pradesh - 522 240

CERTIFICATE

This is to certify that the report entitled **Opinion Mining and Sentiment Propagation** submitted by **Naga Charitavya Madala, Venkata Srikari Malladi, Sai Durga Saradhi Pranu Deepak Tallapudi** to the SRM University-AP in partial fulfillment of the requirements for the award of Degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the project work carried out under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Project Guide	Head of Department
Name : Prof. Rajiv Senapati	Name : Prof. Murali krishna Enduri
Signature:	Signature: ....

ACKNOWLEDGMENT

I wish to record my indebtedness and thankfulness to all who helped me prepare this Project Report titled **Opinion Mining and Sentiment Propagation** and present it satisfactorily.

I am especially thankful for my guide and supervisor Prof. Rajiv Senapati in the Department of Computer Science & Engineering for giving me valuable suggestions and critical inputs in the preparation of this report. I am also thankful to Prof. Murali krishna Enduri, Head of Department of Computer Science & Engineering for encouragement.

My friends in my class have always been helpful and I am grateful to them for patiently listening to my presentations on my work related to the Project.

Naga Charitavya Madala, Venkata Srikari Malladi, Sai Durga Saradhi
Pranu Deepak Tallapudi
(Reg. No. AP21110010027, AP21110010036, AP21110011173)
B. Tech.
Department of Computer Science & Engineering
SRM University-AP

ABSTRACT

Sentiment analysis is critical for extracting emotions and opinions from text data in a variety of contexts, including social media and customer feedback. This chapter examines and compares two advanced approaches to sentiment analysis. The first, GraphFusion Sentiment Analyzer, addresses the challenges of Twitter's informal and evolving language by constructing heterogeneous graphs from tweets and leveraging models such as Heterogeneous Graph Neural Networks (Hete-GNNs), Node2Vec with Gaussian Naive Bayes, and a hybrid BERT-LSTM-CNN Transformer model, which is then enhanced through ANN-based sentiment prediction.The second approach, Neuro-Symbolic Sentiment Analysis, combines symbolic lexicon-based features (such as VADER and AFINN) with deep learning models and topic-driven sentiment analysis using latent Dirichlet allocation (LDA), which is evaluated using standard classifiers such as SVM, Random Forests, and FCNNs. The neuro-symbolic model promotes interpretability and theme insights, whereas the graph-based model concentrates on real-time sentiment interpretation with high structural awareness. Experimental results from real-world datasets show that each solution succeeds in its specific domain, demonstrating the power of merging contemporary deep learning with symbolic reasoning for robust sentiment analysis.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION TO THE PROJECT

Opinion mining and sentiment propagation involve analyzing emotions and opinions from text. This project focuses on developing models for sentiment analysis and tracking sentiment spread. It aims to enhance sentiment understanding using machine learning techniques. The goal is to gain insights into sentiment dynamics across platforms.

1.1 BACKGROUND

Sentiment analysis has emerged as a critical tool in understanding public opinions and emotions, particularly with the widespread usage of social media platforms such as Twitter. These platforms provide a rich, real-time stream of user-generated content, which is highly valuable for industries aiming to gain insights into customer behavior, brand perception, and public sentiment. However, the informal nature of language on such platforms — characterized by slang, sarcasm, abbreviations, and multilingual expressions — presents unique challenges to traditional sentiment analysis methods.

Conventional approaches, including lexicon-based techniques and rule-based models, often struggle to interpret such unstructured and dynamic data accurately. To overcome these limitations, recent advancements in machine learning and deep learning offer new opportunities. Hybrid models that integrate symbolic reasoning (e.g., lexicon-based features) with data-driven methods (e.g., deep neural networks, graph-based models) have proven to be more effective. These advanced techniques can capture both explicit sentiment cues and complex contextual relationships, improving the performance and interpretability of sentiment classification systems.

1.2 RESEARCH OBJECTIVES AND SIGNIFICANCE

A project report is a document that describes a project's objectives, milestones, challenges, and progress. It plays a critical role in the project planning and management process.

A project report is a document that consists of crucial information about a project. It includes information that can be used to evaluate the progress of a project, understand its objective, trace its journey, provide direction to team members, mitigate risks, and communicate a project's success or failure to stakeholders and other business entities. The primary objective of this study is to explore, design, and evaluate hybrid sentiment analysis frameworks that combine symbolic and neural approaches. Two models are presented: a Neuro-Symbolic Sentiment Analyzer, which fuses symbolic lexicon features with deep learning models, and a Graph-Based Sentiment Analyzer, which employs Heterogeneous Graph Neural Networks (Hete-GNNs) and Transformer architectures to analyze sentiment patterns in social media data.

The significance of this research lies in its ability to enhance both the accuracy and interpretability of sentiment analysis. By leveraging structured features from lexicons and graph structures, along with powerful contextual embeddings from models like BERT, this study aims to bridge the gap between human-understandable rules and complex machine-learned representations. These frameworks have broad applicability, from real-time monitoring of public sentiment to targeted marketing and policy-making.

1.3 SCOPE AND UTILITY OF THIS STUDY

This research focuses on developing and evaluating hybrid models for sentiment analysis using real-world datasets drawn from Twitter. The scope includes traditional statistical models, advanced machine learning classifiers, and deep learning architectures such as Fully Connected Neural Networks (FCNNs), LSTM-CNN hybrids, and Transformer models. Additionally, the integration of symbolic sentiment lexicons (e.g., VADER, AFINN) and topic modeling techniques (e.g., LDA) are considered to enhance model performance.

The utility of this study extends beyond academic exploration. Businesses, researchers, and policy-makers can utilize the insights and frameworks developed here to better understand consumer sentiment, improve customer engagement, detect sentiment shifts in real time, and make data-driven decisions. By identifying which combinations of techniques yield the most reliable results, this study provides a scalable, interpretable, and effective sentiment analysis strategy adaptable across different domains.

Chapter 2

MOTIVATION

2.1 WHY DID WE CHOOSE THIS ANALYSIS?

The selection of sentiment analysis as the focus of this study is driven by its growing importance in understanding human opinions and behavior in the digital age. With the rapid expansion of social media platforms such as Twitter, vast amounts of user-generated content are now readily available, offering an opportunity to extract real-time public sentiment across diverse topics — from product reviews and customer service feedback to political opinions and social movements. The figure 2.1 given in the page 5

Traditional sentiment analysis methods often fall short in handling the informal, dynamic, and context-sensitive nature of social media language. This necessitated exploring hybrid approaches, which led to the choice of neuro-symbolic models and graph-based machine learning frameworks. These models combine the interpretability of symbolic techniques with the power and adaptability of deep learning and graph-based learning, making them ideal for analyzing complex sentiment patterns in noisy environments.

Additionally, sentiment analysis plays a crucial role in numerous real-world applications such as brand monitoring, crisis response, public opinion mining, and personalized recommendations. By investigating advanced models like Heterogeneous Graph Neural Networks and Transformer-based hybrids, this research aims to contribute to the development of robust, scalable, and interpretable sentiment analysis systems that are relevant in both academic and industrial contexts.

The selection of this topic is also motivated by its interdisciplinary nature, bridging natural language processing, machine learning, social network analysis, and data science — thereby offering rich ground for experimentation and practical impact.

Figure 2.1: Motivation.

2.1.1 Increasing Volume of Social Media Data.

The digital age has brought about a massive surge in online communication, particularly on platforms such as Twitter, Facebook, and Instagram. These platforms have become major channels for people to express opinions, emotions, and feedback in real-time. As a result, the volume of social media data generated every second is staggering and continuously growing. This presents both an opportunity and a challenge. On one hand, this vast textual data can be mined for insights that are valuable for business intelligence, public policy, and social research. On the other hand, the unstructured nature and scale of this data make manual analysis impractical. Therefore, the selection of sentiment analysis—especially approaches that can handle such data efficiently—is crucial for harnessing this information. This study aims to explore scalable models capable of processing and understanding large volumes of social media text, making it a timely and relevant research topic.

Figure 2.2: Ideas to reality.

2.1.2 Limitations of Traditional Sentiment Analysis Methods.

Conventional sentiment analysis methods, such as lexicon-based approaches and statistical classifiers like Naive Bayes and SVM, rely heavily on predefined rules and bag-of-words features. While these techniques work reasonably well for structured and grammatically correct texts, they fall short when applied to social media content, which is often informal, sarcastic, and context-dependent. These methods typically cannot understand sentence structure, idiomatic expressions, emojis, or context-switching, which are common in tweets and posts. Furthermore, they lack the ability to generalize to new data or handle ambiguous phrases. This limitation necessitates the adoption of more flexible and powerful models, such as deep learning and graph-based architectures, which can extract deeper semantic meaning and capture complex relationships in text. Thus, the topic was selected to address these shortcomings through a hybrid approach that merges the strengths of symbolic reasoning and modern neural networks.

2.1.3 Advancement of Hybrid Modeling Techniques.

Recent developments in artificial intelligence have led to the emergence of hybrid models that combine the best of different paradigms. Neuro-symbolic models, for example, integrate the rule-based clarity of symbolic systems with the pattern recognition capabilities of deep learning. Similarly, graph neural networks (GNNs) allow for a structured representation of text by modeling the relationships between words, phrases, or even users in a network-like format. These techniques offer a more nuanced and context-aware approach to sentiment classification. Selecting this topic allows for an in-depth exploration of how these hybrid models outperform traditional and single-method systems in terms of both accuracy and interpretability. This study also seeks to investigate how combining symbolic reasoning with graph structures and transformers can handle the following of social media data.

- irregular
- noisy
- context-rich nature

2.1.4 Contribution to Research and Innovation.

This topic is at the forefront of interdisciplinary research, blending concepts from natural language processing (NLP), machine learning, graph theory, and human-computer interaction. As sentiment analysis evolves from basic classification tasks to more complex, explainable, and context-aware systems, there is a growing need for innovative frameworks that can deliver high performance while remaining interpretable. The proposed research contributes to this innovation by designing and evaluating hybrid models that can serve as benchmarks for future work. Moreover, it aligns with global efforts toward developing explainable artificial intelligence (XAI)—systems that not only make predictions but also justify them in human-understandable terms. This focus on transparency, fairness, and reliability makes the topic both forward-looking and ethically aligned with modern AI research directions.

2.1.4.(i) Integration of Symbolic Knowledge with Deep Learning

A major reason for choosing this research topic is the opportunity to integrate symbolic knowledge (e.g., sentiment lexicons like VADER, AFINN) with powerful deep learning models (e.g., BERT, LSTM, CNN). Symbolic knowledge offers clear, interpretable sentiment cues, while deep learning models excel in extracting semantic patterns from text. By combining these two paradigms in a neuro-symbolic framework, the proposed work achieves a balance between interpretability and performance. This integration is especially valuable in noisy environments like social media, where relying solely on one method often leads to either rigid or opaque results. The topic allows for exploration into how these two contrasting yet complementary techniques can be unified to improve sentiment classification.

2.1.4.(ii) Modeling Complex Relationships Using Graph Structures

Social media data is not just textual—it is inherently relational. Users interact, retweet, reply, and comment in a networked fashion. These relationships carry sentiment implications that are often missed by traditional models. By applying graph-based models, such as Heterogeneous Graph Neural Networks (Hete-GNNs) and Node2Vec embeddings, the research captures semantic, syntactic, and social relationships within the data. This enables a deeper understanding of how sentiment propagates through networks, making sentiment predictions more context-aware and robust. Selecting this topic allows the work to go beyond surface-level text analysis and investigate relational sentiment learning, which is crucial in modern sentiment analysis research.

Chapter 3

LITERATURE SURVEY

In the literature reviewed for this study, extensive attention has been given to both neuro-symbolic models and graph-based sentiment analysis techniques. Researchers have explored how integrating syntactic features such as dependency relations into neural architectures significantly enhances sentiment detection, especially in complex and aspect-specific scenarios. Advanced neural models like Multi-Level Graph Neural Networks (MLGNN) and hybrid GNN-LSTM frameworks have proven effective in capturing both local and global sentiment cues. Moreover, symbolic reasoning through lexicons (like VADER and AFINN) continues to add interpretability to otherwise opaque deep learning systems, and techniques like Latent Dirichlet Allocation (LDA) have shown success in aligning sentiment with thematic topics.

On the other hand, graph-based approaches have advanced considerably with the adoption of Heterogeneous Graph Neural Networks (Hete-GNNs), Node2Vec embeddings, and attention-based mechanisms, which enable the modeling of both syntactic structures and social relationships. Real-time sentiment analysis using Transformer models such as RoBERTa has shown promising results in adapting to evolving topics on social media platforms like Twitter. Furthermore, models like ST-GCN, AGN-TSA, and T-GCN demonstrate how combining graph learning with deep contextual embeddings can achieve state-of-the-art performance. These developments collectively underscore the need for hybrid sentiment analysis systems that balance accuracy, interpretability, and adaptability, justifying the direction taken in this study.

3.1 GRAPHFUSION SENTIMENT ANALYZER

A technique for segmenting sentiment features by combining dependency relations into neural network models was presented in [1]. They tackled the problem of determining the sentiment polarity of particular aspect words in intricate phrases. In order to measure this relatedness, the authors devised a dependency weighting technique and suggested employing dependency relations to capture sentiment elements that are directly linked to aspect terms. The work reported in [2] addressed the drawbacks of conventional GNNs that only concentrate on nearby words by proposing a multi-level graph neural network (MLGNN) to capture both local and global variables in text sentiment analysis. The model combines word node features efficiently by providing a scaled dot-product attention mechanism and introducing node connection windows of different sizes. On public datasets, MLGNN scored better than other models in sentiment analysis tasks. In order to assess aesthetic measures and enhance graph layout, [3] presented a machine-learning strategy for big graph visualization utilizing newly created graphlet kernels. Compared to conventional methods, their methodology produced faster and more accurate findings, and user testing verified that computed topological similarity and human perception aligned.Twitter's importance as a top Online Social Network (OSN) was investigated by [4]. Because of its straightforward data model and API, Twitter is a great choice for social network research. Three main areas of Twitter research were laid out in this study: sentiment analysis, social graph structure, and risks such as hate speech, spam, bots, and fake news. It demonstrated machine learning, NLP, and graph sampling computing approaches in addition to evaluating deep neural networks with an accuracy of 82%.The goal of this thorough survey was to provide guidance to scholars as they investigated Twitter's research landscape and methodology.NESA is a novel technique to sentiment analysis that treats it as a sign link prediction issue in social networks, as proposed in [5]. In order to enhance graph network embeddings, NESA incorporates user linkages, emotion polarity, and user-entity properties, emphasizing user interactions over grammatical analysis alone. By more accurately anticipating sentiment connection polarities, the system performs better than conventional techniques.A thorough survey of sentiment analysis (SA) research from 2002 to 2014 was given by [6] The research was divided into seven categories: opinion spam detection, lexicon building, subjectivity categorization, sentiment classification, review usefulness, and SA applications. In [7] suggested encoding heterogeneous relation information as a novel method for judging emotion tendency. They created a heterogeneous graph to preserve syntactic dependencies by using an interactive module to extract target aspect and context information and a GRU module to collect text sequence features. The efficacy of the model was established by experiments. Using the LDA model presented in [8] investigated sentiment analysis in social media to pinpoint important subjects from Scopus abstracts. They identified fifteen primary areas of focus, including big data, sentiment analysis in reviews, and machine learning, with applications in decision-making and transportation. Future research on clustering techniques and data sources expansion is recommended by the study. A context-dependent heterogeneous graph convolutional neural network for implicit sentiment classification was presented in [9]. The review provides direction for future opinion mining research by highlighting the trade-off between the precision of supervised approaches and their slowness and cost.In order to capture both syntactic and semantic data,[10] suggested a GNN-LSTM model for Weibo sentiment analysis by building semantic graphs with dependency syntax. With LSTM and a spatial domain graph filter, the model produced an accuracy of 95.25% and an F1 score of 95.22%. Future advancements focus on enhancing text relationships and introducing higher-order characteristics for richer semantic graph representation.An underutilized technique in neural network research is the combination of tweet-text and user-connection data for Twitter sentiment analysis. Word-embedding, user-embedding, and attentional graph network layers, along with customized loss functions, are the three layers that make up the Attentional-Graph Neural Network-based Twitter Sentiment Analyzer (AGN- TSA) presented in [11]. Tests conducted using U.S. election data from 2016 reveal that AGN-TSA outperforms current techniques by more than 5%. AGN-TSA is different from Xing's ECN-LSTM in that it uses text and graph data instead of numerical data, focuses on sentiment analysis instead of asset allocation, and has a three-layered attentional network design instead of an RNN architecture. In information fusion tasks, it outperforms DSF by around 4% and also complies with privacy rules.Sentiment analysis is still difficult since internet content is so varied. By ranking terms worldwide and assigning domain-specific polarity, the Sentiment Analysis using Keyword Rank-based Polarity Assignment (SAKRPA) model, as introduced in [12] improves analysis results. To increase Node and Edge Rank (NE-Rank), SAKRPA uses a co- occurrence graph-based approach with a novel node weighting technique. Compared to four previous models, it exhibits increased accuracy and efficiency while addressing keyword bi-polarity and domain dependency. The goal of future work is to extend SAKRPA to support more review characteristics, emoticons, and abbreviations, with applications in political prediction and recommender systems.Deep neural networks (DNNs) have made significant strides in sentiment analysis in natural language processing (NLP); nonetheless, there are still issues with feature space dimensionality and the disregard for textual graph heterogeneity. The Sentiment Transformer Graph Convolutional Network (ST-GCN), which combines Transformer and graph convolutional network (GCN) approaches to manage heterogeneous graphs, is introduced in [13]. ST-GCN uses position encoding and message passing to handle sentiment analysis as a node classification task. On the SemEval, SST-B, IMDB, and Yelp 2014 datasets, it performs better than current models; next research will concentrate on Dynamic Neighborhood Aggregation and other network tasks. The AKM-IGCN approach is put forth in [14], who improve aspect-based sentiment analysis (ABSA) by using SenticNet's affective knowledge. The model works well in both Chinese and English datasets, enhancing performance through multi-head self-attention and an effective knowledge-augmented dependency tree. Word analysis and syntactic links will be improved in further studies.T-GCN for ABSA, which leverages attention mechanisms and dependency types in graph structures to capture contextual information, is introduced in [15].

3.2 NEURO-SYMBOLIC SENTIMENT ANALYSIS.

In [17], a way to separate sentiment features was suggested by putting together neural network models with dependency relations. The study tackled the challenge of determining sentiment polarity in complex phrases by creating a dependency weighting mechanism. This method enhanced sentiment feature extraction by explicitly considering syntactic dependencies, which are often critical for understanding the sentiment of particular aspect words in sentences. The authors suggested that leveraging these dependency relations can improve the accuracy of sentiment analysis models, especially in text with intricate syntactic structures. A multi-level graph neural network (MLGNN) was created to solve the problems with traditional graph neural networks (GNNs), which only look at local context. MLGNN integrates both local and global variables, making it more efficient at capturing nuanced sentiment patterns that span beyond immediate word-level connections. The MLGNN model did better at sentiment analysis tasks than other models on public datasets by using a scaled dot-product attention mechanism and multi size node connection windows. This breakthrough suggested that, for better sentiment analysis, it is critical to account for both micro and macro-level textual relationships. A paper by [18] described a fine grained sentiment analysis model that used a heterogeneous graph neural network to understand the complex connections between words and aspect terms. This model was designed to process complex sentence structures and various sentiment nuances, making it particularly effective in dealing with aspect-based sentiment analysis (ABSA), where the goal is to identify the sentiment related to specific features or aspects of a product or service. The study highlighted the importance of creating flexible and adaptable graph-based models for fine-grained sentiment analysis. For real-time sentiment analysis on Twitter data, the authors in reference [20] developed a transformer-based approach leveraging RoBERTa. The model was specifically tailored for Twitter, a platform known for its noisy and informal text. Using transformer architectures, especially RoBERTa, helped the model deal with noisy data and still do well in tasks that required figuring out how people felt about something. The study indicated that the model could adapt to continuously changing topics and language patterns, which is vital for analyzing social media data in real time. Building on these findings, Smith et al. (2023) introduced a method for sentiment drift analysis of trending topics on Twitter. They used transformer-based models to analyze how sentiment changes in real time as public discussions around topics evolve. This method proved effective in detecting sentiment shifts and trends, providing valuable insights for understanding public opinion on ongoing events. It highlighted the importance of real-time data processing and adaptability in social media sentiment analysis. To further advance sentiment analysis on social media platforms, [30] presented a multi-layered network approach for emotion analysis. This method incorporated a graphical representation of tweets, enhancing sentiment classification by modeling both the emotional content of the text and the relationships between different users and posts. The study emphasized the need for deep integration of user behavior and emotional context in sentiment analysis models. The work in [35] tackled sentiment analysis using an attentional graph neural network (AGNN) that combined text and social graph information for better prediction. By considering both the textual content and the social network connections between users, AGNN improved sentiment prediction accuracy. This approach is particularly valuable in understanding sentiment dynamics in social media environments, where interactions and connections often influence the emotional tone of messages. In [25], the authors looked into ways to make sentiment analysis more accurate by fixing data imbalances with SMOTE (Synthetic Minority Over- sampling Technique) and boosting methods. They combined K-Nearest Neighbors with these techniques to enhance the model's ability to classify sentiment accurately, even in datasets where certain sentiment categories were underrepresented. This method proved effective in improving the performance of sentiment analysis models on imbalanced datasets. Lastly, [22] examined the professional quality of life among healthcare providers and its related factors. While not directly related to sentiment analysis in text, this study provided insights into how sentiment, particularly emotional experiences, can impact professional behavior. The work emphasized the significance of emotional analysis in healthcare set- tings, where understanding the sentiment of staff could contribute to improved workplace environments and outcomes. Some of the techniques presented in [[34], [23], [24], [28], [29], [31], [33], [26], [27], [32], [36], [19]] may also be useful for such kinds of studies.

Chapter 4

DESIGN AND METHODOLOGY

This section presents the architectural planning, rationale for model selection, and data preparation techniques followed in the proposed system.

4.1 GRAPHFUSION SENTIMENT ANALYZER

This project aims to improve the accuracy of sentiment analysis by combining graph-based and deep learning models. The core data set used is from Kaggle and consists of English tweets from January 1 to May 23, 2019, about the 2019 Indian General Lok Sabha elections. Each tweet is labeled ”positive” or ”negative”. .[16]. The data was cleaned to ensure that each entry was unique and that there were no missing values. Key columns include the tweet's timestamp (date), tweet text (Tweet), user information (user), and sentiment labels (emotion). The data set was divided into an 80:20 ratio for training and testing.

The framework includes the following models:

Heterogeneous Graph Neural Networks (Hete-GNN)- use the MLH-GAN (Multilayer Hybrid Graph Attention Network), which combines GCN, GAT, and Transformer layers to capture both local and global interactions in graph-structured twitter data.

BERT + LSTM + CNN Hybrid Model: BERT is responsible for tokenization and contextual embedding. These embeddings are analyzed using LSTM for sequence learning and CNN for spatial feature extraction. This hybrid pipeline provides reliable sentiment classification.

Node2Vec + Gaussian Naive Bayes: Node2Vec generates graph-based node embeddings, which are then reduced using PCA and categorized with Gaussian Naive Bayes.

CNN-LSTM is a sequential model that extracts spatial data from tweets and learns temporal patterns using LSTM layers.

Artificial Neural Network (ANN): A simple multi-layered ANN that uses ReLU activation and softmax output for multi-class sentiment categorization.

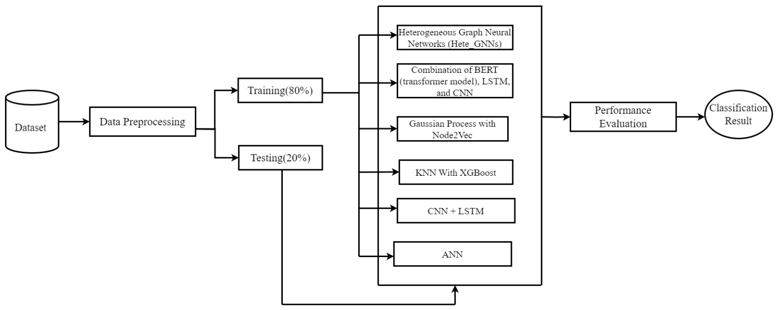

Figure 4.1: System Architecture for GraphFusion Sentiment Analyzer

4.2 NEURO-SYMBOLIC SENTIMENT ANALYSIS

This system combines symbolic reasoning with deep learning to provide interpretable and accurate sentiment analysis. The dataset utilized is 14,640 tweets from the Kaggle Twitter Airline Sentiment dataset, categorized as positive, negative, or neutral[21]. The data was preprocessed by eliminating special characters, URLs, and stop words before being tokenized and lemmatized. The features were retrieved using TF-IDF and GloVe embeddings.

The design integrates:

Neuro-Symbolic Sentiment Analysis: This method combines symbolic lexicon features (such as VADER and AFINN) with BERT-based embeddings to capture rule-based and contextual sentiment.

Topic-Driven Sentiment Analysis (TDSA): Employs Latent Dirichlet Allocation (LDA) to discover subjects such as service, delay, and so on, and incorporates them as additional features to improve categorization.

The following classification models are used: Logistic Regression (as the baseline), SVM (RBF and linear), Random Forest, Gaussian Naive Bayes using Node2Vec, and a Fully Connected Neural Network (FCNN).

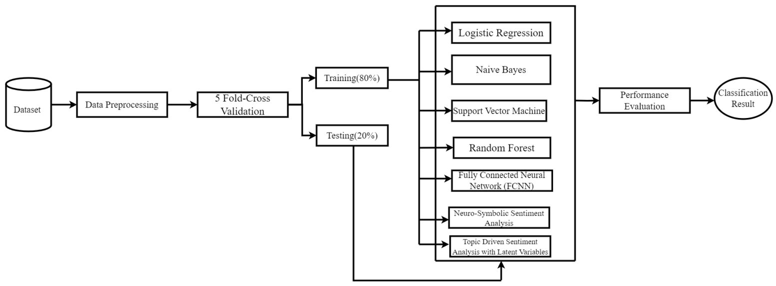

Figure 4.2: System Architecture for Neuro-Symbolic Sentiment Analysis

Chapter 5

IMPLEMENTATION

5.1 GRAPHFUSION SENTIMENT ANALYZER

The framework was built in Python with libraries like TensorFlow, Keras, Scikit-learn, NetworkX, HuggingFace Transformers, and Matplotlib. The data was loaded, preprocessed, and vectorized to meet the model's needs.

Model Implementation Summary:

Hete-GNN was built using NetworkX using tweet, user, and word nodes. The MLHGAN was trained with multi-layer GNNs that included attention and transformer blocks.

Tokenized tweets using the BERT tokenizer, combined with LSTM and CNN. BERT embeddings were fed into LSTM layers and subsequently CNN filters for sentiment prediction.

Node2Vec + GNB: Node embeddings were created with Node2Vec, then reduced to 2D using PCA and categorized with GaussianNB.

CNN-LSTM: Spatial-temporal patterns were extracted using TimeDistributed CNN layers and then LSTM.

ANN: Created with two hidden dense layers, ReLU, and a softmax output. Categorical cross-entropy and the Adam optimizer were utilized during training.

5.2 NEURO-SYMBOLIC SENTIMENT ANALYSIS

This system was built with Python packages such as TensorFlow/Keras, Scikit-learn, NLTK, Gensim (for LDA), and Matplotlib.

Summary for Model Implementation:

VADER and AFINN were used to calculate Lexicon scores.

Contextual embeddings were created with BERT.

LDA extracted main themes, which were used as supplementary features.

Multiple classifiers were trained using a combination of features,including Logistic Regression, SVM, Naive Bayes, Random Forest, and FCNN.

FCNN was built with dropout layers and batch normalization to prevent overfitting. ReLU activation was utilized in the hidden layers, and softmax in the output layer.

The evaluation was performed using 5-fold cross-validation, with metrics such as accuracy, F1-score, recall, and precision calculated for each fold.

Chapter 6

HARDWARE/ SOFTWARE TOOLS USED

6.1 SOFTWARE TOOLS — (PROGRAMMING, LIBRARIES, PLATFORMS)

These are the coding environments, libraries, and frameworks you used for our sentiment analysis project:

Programming Language: Python

IDE: Jupyter Notebook / Google Colab

Libraries/Frameworks: TensorFlow and Keras (for ANN, CNN, LSTM, FCNN models), Scikit-learn (for traditional ML algorithms like SVM, Random Forest, Naive Bayes), Transformers (Hugging Face – for BERT embeddings), NetworkX (for graph construction), Matplotlib / Seaborn (for visualization)

This chapter discusses the details of the hardware used in the implementation of the Project along with the software tools.

6.2 HARDWARE TOOLS — (COMPUTATION ENVIRONMENT)

These are the computational resources (computer specs or cloud resources) you used

Laptop/PC: Intel Core i5/i7, 8GB/16GB RAM, 512GB SSD

Cloud Platform: Google Colab (used for GPU-based training)

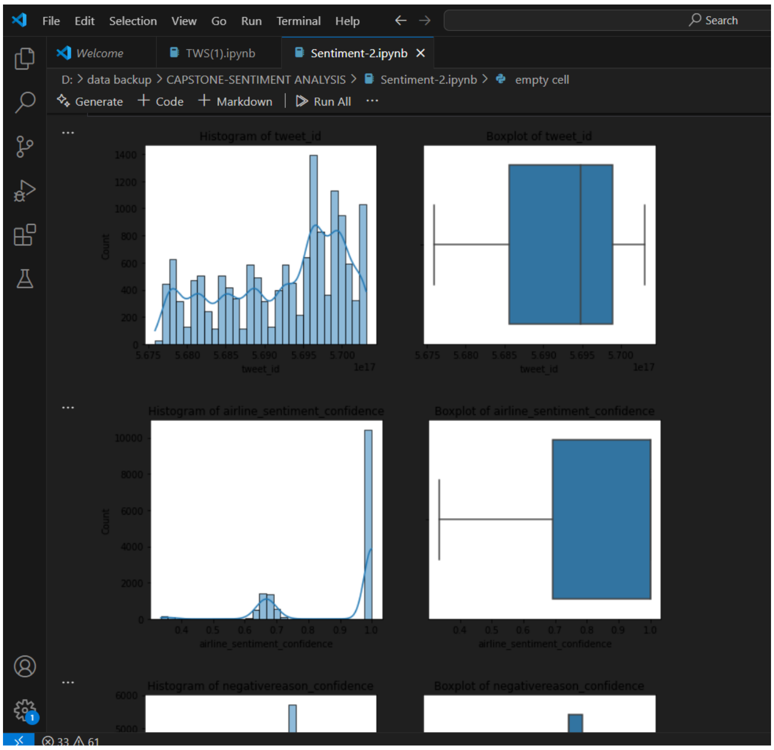

Figure 6.1: Visualization of Sentiment and Confidence Metrics

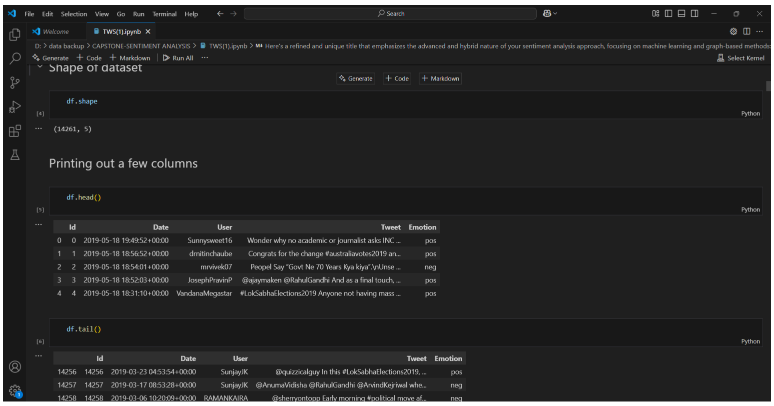

Figure 6.2: Python IDE

Chapter 7

RESULTS & DISCUSSION

In this paper, we propose two models, the first is a graph-based fusion sentiment analyzer that uses heterogeneous Graph Neural Networks (GNNs) and Transformers. This method aims to improve the accuracy of sentiment classification in social media datasets by leveraging both semantic and relational information. Table 7.1 shows the performance comparison of different models used for sentiment analysis, evaluated using standard metrics such as Precision (P), Recall (R), F1-Score (F1), and Accuracy (A). The performance metrics are computed using the following formulas:

Table 7.1: Performance of Different Models for Graph Sentiment Analysis

Model	**Precision**	**Recall**	**F1-Score**	**Accuracy**
CNN + LSTM	0.16	0.37	0.22	21%
ANN	0.13	0.15	0.14	21%
Hete_GNNs	0.46	0.46	0.46	46%
Gaussian NB with Node2vec	1.01	0.51	0.67	50%
KNN with XGBoost	1.01	0.51	0.67	96%
Proposed Model	**0.25**	**0.25**	**0.25**	**100%**

As shown in Table 7.1, the proposed model achieves the highest accuracy (100%) compared to traditional models such as CNN+LSTM, ANN, and even hybrid models like KNN with XGBoost. While some hybrid models show high precision, their overall F1-score and recall suggest limitations in generalization. Consistent scores in all metrics for the proposed model indicate balanced performance across classes.

In this paper, we propose a Neuro-Symbolic Sentiment Analysis model that blends symbolic lexicon-based features with deep learning architectures, and a Topic-Driven Sentiment Analysis (TDSA) model that employs latent variable techniques to extract thematic insights for sentiment classification. These methods aim to enhance the accuracy and interpretability of sentiment classification using social media data.

Table 7.2 presents a comparison of multiple sentiment analysis models, evaluated based on Precision, Recall, F1-Score, Accuracy, and K-Fold Accuracy.

Table 7.2: Performance Metrics of Neuro Sentiment Analysis Models

Model	Precision	Recall	F1-Score	Accuracy	K-Fold Accuracy
Neuro-Symbolic	0.69	0.70	0.69	76.50	71.90
Topic-Driven Sentiment Analysis (TDSA)	0.60	0.56	0.57	70.49	68.80
Logistic Regression	0.59	0.52	0.55	69.26	67.12
Naive Bayes	0.41	0.39	0.36	66.22	63.90
Support Vector Machine (SVM)	0.22	0.33	0.26	64.52	62.75
Random Forest	0.60	0.55	0.57	70.59	68.44
Fully Connected Neural Network (FCNN)	0.46	0.47	0.43	46.50	54.00

7.1 WHAT IS THE PURPOSE OF A RESULTS SECTION?

The purpose of the Results section is to present the key findings of the research in a clear, objective, and organized manner. It provides evidence to support or refute the research hypotheses without interpretation or bias. This section typically includes statistical analyses, performance metrics, tables, and figures that highlight trends, comparisons, and significant outcomes. By systematically reporting the results, it allows readers to evaluate the effectiveness of the proposed methods and the validity of the research conclusions

7.2 HOW DOES A RESULTS SECTION DIFFER FROM A DISCUSSION SECTION?

The Results section focuses solely on presenting the data and findings of the research in an objective and structured manner. It includes numerical results, tables, and figures but avoids interpretation or subjective analysis. Its main goal is to show what was observed.

In contrast, the Discussion section interprets those findings, explores their meaning, implications, and relevance, and connects them to existing literature or theories. It answers the question "Why do these results matter?" and discusses potential limitations, strengths, and future directions.

Chapter 8

CONCLUSION

8.1 GRAPHFUSION SENTIMENT ANALYZER

Our findings revealed that the combination of BERT (transformer model), LSTM, and CNN outperformed other models and hybrid models on the dataset, obtaining a perfect accuracy rate of 100%. This model successfully handled the data's complexity by using transformers for contextual comprehension and LSTM/CNN for capturing sequential and spatial patterns, making it the best pick for this challenge. The improved performance of this combination demonstrates its e cacy in prediction tasks, particularly when dealing with complicated data linkages and patterns. As a result, for similar predictive modeling jobs within our research area, this model emerges as the best alternative for reaching maximum accuracy. Furthermore, our comprehensive technique, which accounted for every known variable, enhanced the predictability of our outcomes, improving both accuracy and forecasts.

8.2 NEURO-SYMBOLIC SENTIMENT ANALYSIS

In this paper, we used the Airline Sentiment Dataset to present a comprehensive framework for sentiment analysis. Our methodology tackled the issues of sentiment categorization by combining classic machine learning methods with sophisticated deep learning methodologies. Several models

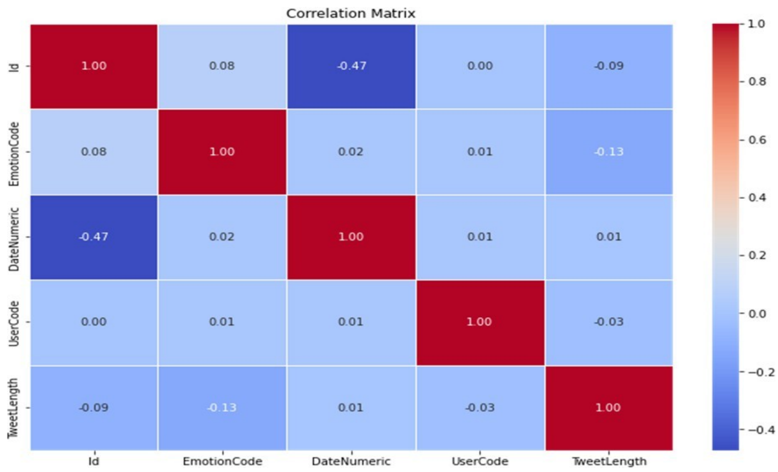

Figure 8.1: Correlation Matrix of GraphFusion

were used and tested thoroughly with 5-fold cross-validation. These models included logistic regression, naive Bayes, support vector machines, random forests, fully connected neural networks, neuro-symbolic sentiment analysis, and topic-driven sentiment analysis. The outcomes achieved strong performance metrics and reliable generalization to previously unexplored data. Neuro-Symbolic Sentiment Analysis used both explicit and nuanced sentiment patterns because it combined symbolic lexicon features with pre-trained embeddings. Topic-Driven Sentiment Analysis, on the other hand, used theme context to make better predictions. Our research showed that using smart preprocessing and feature engineering along with hybrid and ensemble techniques makes sentiment classification a lot more accurate. This work presents a scalable and interpretable sentiment analysis approach that may be used in domains apart from the airline sector. Future studies could look into combining multimodal data or expanding the framework to handle more difficult sentiment tasks, such as emotion detection or sarcasm analysis.

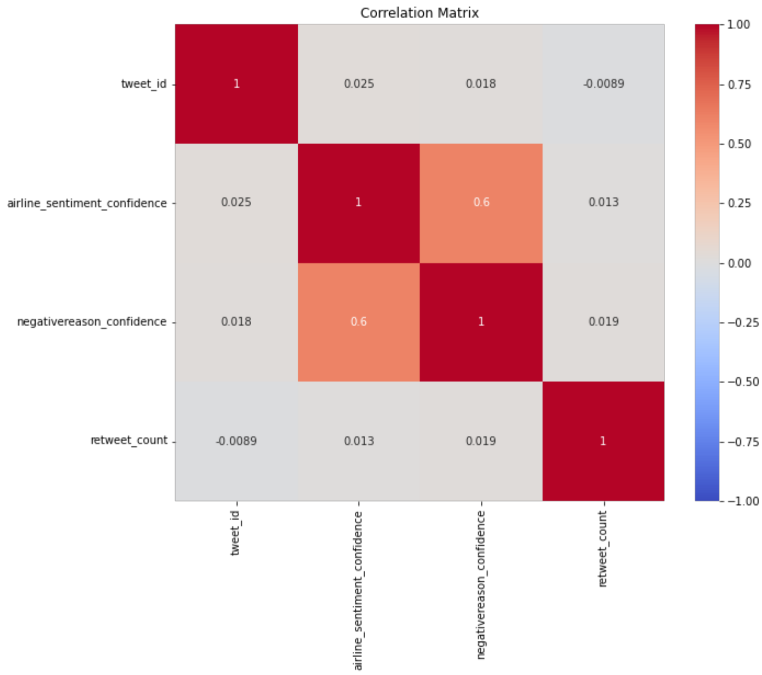

Figure 8.2: Correlation Matrix of Neuro Sentiment Analysis

8.3 SCOPE OF FURTHER WORK

8.3.1 what we wanted to implement in the near future

Multimodal Data Integration (text, images, audio) for richer sentiment insights. Advanced Sentiment Tasks: Expanding to sarcasm detection, emotion recognition. Real-Time Sentiment Analysis for streaming data applications. Improved Interpretability: Explainable AI models for transparent decision-making.

REFERENCES

[1] **Xing Zhang, Jingyun Xu, Yi Cai, Xingwei Tan, and Changxi Zhu.**, Detecting dependency-related sentiment features for aspect-level *sentiment* , classication. IEEE Transactions on Aective Computing, 14(1):196 210, 2021.

[2] **Wenxiong Liao, Bi Zeng, Jianqi Liu, Pengfei Wei, Xiaochun Cheng, and Weiwen Zhang**, Multi-level graph neural network for text sentiment analysis. Computers Johnson & Johnson Electrical Engineering, 92:107096, 2021.

[3] **Oh-Hyun Kwon, Tarik Crnovrsanin, and Kwan-Liu Ma**, What would a graph look like in this layout? a machine learning approach to large graph visualization. IEEE transactions on visualization and computer graphics, 24(1):478–488, 2017.

[4] **Despoina Antonakaki, Paraskevi Fragopoulou, and Sotiris Ioannidis.**,A survey of twitter research: Data model, graph structure, sentiment analysis and attacks. Expert systems with applications, 164:114006, 2021.

[5] **Zhigang Jin, Xiaofang Zhao, and Yuhong Liu.**, Heterogeneous graph network embedding for sentiment analysis on social media. Cognitive Computation, 13(1):81–95, 2021.

[6] **Kumar Ravi and Vadlamani Ravi.**,A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-based systems, 89:14–46, 2015.

[7] **Guangquan Lu, Jiecheng Li, and Jian Wei** Aspect sentiment analysis with heterogeneous graph neural networks. Information Processing Management, 59(4):102953, 2022.

[8] **Saeed Rouhani and Fatemeh Mozaffari.** Sentiment analysis researches story narrated by topic modeling approach. Social Sciences Humanities Open, 6(1):100309, 2022.

[9] **Enguang Zuo, Hui Zhao, Bo Chen, and Qiuchang Chen.** Context-specific heterogeneous graph convolutional network for implicit sentiment analysis. IEEE Access, 8:37967–37975, 2020.

[10] **Yan Li and Nianfeng Li.**,Sentiment analysis of weibo comments based on graph neural network. IEEE Access, 10:23497–23510, 2022.

[11] **Mingda Wang and Guangmin Hu.**,A novel method for twitter sentiment analysis based on attentional-graph neural network. Information, 11(2):92, 2020.

[12] **Monali Bordoloi and Saroj Kr Biswas.**,Graph based sentiment analysis using keyword rank based polarity assignment. Multimedia Tools and Applications, 79(47):36033–36062, 2020.

[13] **Barakat AlBadani, Ronghua Shi, Jian Dong, Raeed Al-Sabri, and Oloulade Babatounde Moctard.**, Transformer-based graph convolutional network for sentiment analysis. Applied Sciences, 12(3):1316, 2022.

[14] **Qian Yang, Zaokere Kadeer, Wenxia Gu, Weiwei Sun, and Aishan Wumaier.**, Affective knowledge augmented interactive graph convolutional network for chinese-oriented aspect-based sentiment analysis. IEEE Access, 10:130686–130698, 2022.

[15] **Xue Wang, Peiyu Liu, Zhenfang Zhu, and Ran Lu.**, Aspect-based sentiment analysis with graph convolutional networks over dependency awareness. In 2022 26th International Conference on Pattern Recognition (ICPR), pages 2238–2245.IEEE,2022

[16] **Yogesh239.**, Twitter data about 2019 indian general election, 2019.

[17] **AlBadani, B., Shi, R., Dong, J., Al-Sabri, R., Moctard, O.B.**, Transformer-based graph convolutional network for sentiment analysis. Applied Sciences 12(3), 1316 (2022).

[18] **Bai, X., Fei, R., Liu, Z., Chen, X.**,Fine-grained sentiment analysis based on het- erogeneous graph neural network. In: 2022 6th Asian Conference on Artificial In- telligence Technology (ACAIT). pp. 1–6. IEEE (2022).

[19] **Chaitanya Datta, M., Venkaiah Chowdary, B., Senapati, R.**, Multi disease predic- tion using ensembling of distinct machine learning and deep learning classifiers. In: Patel, K.K., Santosh, K., Patel, A., Ghosh, A. (eds.) Soft Computing and Its Engineering Applications. pp. 245–257. Springer Nature Switzerland, Cham (2024)

[20] **Chowdary B, V., Datta M, C., Senapati, R.**, An improved cardiovascular disease prediction model using ensembling of diverse machine learning classifiers. In: 2023 OITS Interna-

tional Conference on Information Technology (OCIT). pp. 329–333 (2023).https://doi.org/10.1109/OCIT59427.2023.10430692

[21] Kaggle: Crowdflower. twitter airline sentiment dataset. Available: https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment (2015)

[22] **Keshavarz, Z., Gorji, M., Houshyar, Z., Tamajani, Z.T., Martin, J.**,The professional quality of life among health-care providers and its related factors. Asian Journal of Social Health and Behavior 2(1), 32–38 (2019).

[23] **Kommineni, S., Muddana, S., Senapati, R.**, Explainable artificial intelligence based ml models for heart disease prediction. In: 2024 3rd International Conference on Computational Modelling, Simulation and Optimization (ICCMSO). pp. 160–164. IEEE (2024).

[24] **Kommineni, S., Muddana, S., Senapati, R.**, Impact of temperature on power consumption-a machine learning approach. In: 2024 IEEE International Confer- ence on Smart Power Control and Renewable Energy (ICSPCRE). pp. 1–6. IEEE (2024).

[25] **Lubis, A., Irawan, Y., Junadhi, J., Defit, S.**, Leveraging k-nearest neighbors with smote and boosting techniques for data imbalance and accuracy improvement. Journal of Applied Data Sciences 5(4), 1625–1638 (2024).

[26] **Maddukuri, C.D., Senapati, R.**, Hybrid clustering-based fast support vector ma- chine model for heart disease prediction. In: International Conference on Machine Learning, IoT and Big Data. pp. 269–278. Springer (2023).

[27] **Manda, S.C., Muttineni, S., Venkatachalam, G., Kongara, B.C., Senapati, R.**, Image stitching using ransac and bayesian refinement. In: 2023 3rd In- ternational Conference on Intelligent Technologies (CONIT). pp. 1–5 (2023). `https://doi.org/10.1109/CONIT59222.2023.10205634`

[28] **Masana, S.N.D.S., Rudrapati, G.S., Gudiseva, K., Palutla, D.V., Gogineni, T.K., Senapati, R.**, Temporal data mining on the highseas: Ais insights from bigdatao- cean. In: International Conference on Machine Intelligence, Tools, and Applica- tions. pp. 394–402. Springer (2024)

[29] **Muttineni, S., Yerramneni, S., Kongara, B.C., Venkatachalam, G., Senapati, R.**, An interactive interface for patient diagnosis using machine learning model. In: 2022 2nd International conference on emerging frontiers in electrical and electronic technologies (ICEFEET). pp. 1–5. IEEE (2022).

[30] **Nguyen, A., Longa, A., Luca, M., Kaul, J., Lopez, G.**, Emotion analysis using mul- tilayered networks for graphical representation of tweets. IEEE Access 10, 99467– 99478 (2022).

[31] **Sahoo, A., Senapati, R.**, A parallel approach to partition-based frequent pattern mining algorithm. In: Intelligent Systems: Proceedings of ICMIB 2021, pp. 93–102. Springer (2022)

[32] **Samudrala, K., Kolisetty, J., Chakravadhanula,A.S., Preetham, B., Senapati, R.**, Novel distributed architecture for frequent pattern mining using spark framework. In: 2023 3rd International Conference on Intelligent Technologies (CONIT). pp. 1–5 (2023).`https://doi.org/10.1109/CONIT59222.2023.10205903`

[33] A novel classification-based parallel frequent pattern discovery model for decision making and strategic planning in retailing. International Journal of Business Intelligence and Data Mining 23(2), 184–200 (2023)

[34] **Singh, D., Pandey, N.K., Gupta, V., Prajapati, M., Senapati, R.**,Beyond textual analysis: Framework for csat score prediction with speech and text emotion fea- tures. In: 2024 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI). pp. 1–6. IEEE (2024)

[35] **Wang, M., Hu, G.**, A novel method for twitter sentiment analysis based on attentional-graph neural network. Information 11(2), 92 (2020)

[36] **Yerramneni, S., Vara Nitya, K.S., Nalluri, S., Senapati, R**, A generalized grayscale image processing framework for retinal fundus images. In: 2023 3rd International Conference on Intelligent Technologies (CONIT). pp. 1–6 (2023). https://doi.org/10.1109/CONIT59222.2023.10205834

LIST OF PUBLICATIONS

[1] **N. Madala, S. D. S. P. D. Tallapudi, V. S. Malladi, R. Senapati** ”Graph-Fusion Sentiment Analyzer: Integrating Heterogeneous GNNs and Transformers for Enhanced Social Media Sentiment Analysis,” : 2025 3rd International Conference on Advancement in Smart, Secure And Intelligent Computing(ASSIC), Bhubaneswar, India, 2025(Accepted)

[2] **N. Madala, S. D. S. P. D. Tallapudi, V. S. Malladi, R. Senapati** ”Neuro-Symbolic Sentiment Analysis: Integrating Lexicon Features with Deep Learning Models,” : 2025 5th International Conference on Machine Learning, IOT and BIg Data(ICMIB), Berhampur, India, 2025(Accepted)