Enhancing Sentiment Analysis With Emotion And Sarcasm Detection: A Transformer-Based Approach

Mr. Suryavamshi Sandeep Babu[1], S.V. Suryanarayana[2], M. Sruthi[3], P. BhagyaLakshmi[4], T.Sravanthi[5], M.Spandana[6]

[1.] Assistant Professor, Department of CSE (Artificial Intelligence and Machine Learning)
[2]Professor, Computer Science Engineering (Data Science), CVR College of Engineering, Telangana-501510, India
[(3,4,5,6)] B. Tech IV Year Students, Department of CSE (Artificial Intelligence and Machine Learning),
[(1,3,4,5,6)]Vignan's Institute of Management and Technology for Women, Hyderabad, Telangana-501301, India [1]s.sandeep.0001@gmail.com, [2] suryahcu@gmail.com, [3] shruthimangu01@gmail.com,
[4] bhagyalaxmipolu6@gmail.com, [5] thagaramsravanthi@gmail.com, [6]spandanan71@gmail.com

Abstract

Our research focuses on analyzing the reviews generated on ecommerce sites like Amazon, which is complex due to the diverse ways in which the customers express themselves. They may be informal, sarcastic, and contain underlying and hidden sentiments making it difficult to analyze the sentiments meticulously. Traditional methods have been successful in training sentiment analysis models which predict whether a given statement or review is positive, negative or neutral. They fail to record deeper and underlying emotions like frustration, joy, anger and also, cannot detect sarcasm, where a positively shaped statement is actually negative. Therefore, our research puts forward an advanced sentiment analysis system that incorporates both Natural Language Processing (NLP) and Machine Learning techniques to bridge these gaps. Consequently, integrating sentiment analysis, emotion and sarcasm detection gives a better understanding of the product and customer feedback. To enhance the readability, the system also consists of visualizations of the sentiments and emotions through interactive pie charts and word clouds. This hence helps businesses to take data-centric decisions, and also help customers get a concise analysis of the product. This research emphasizes on the usage of VADER and BERT for sentiment analysis. Additionally, a neural network is used for sarcasm detection and Amazon reviews are scraped using Beautiful Soup.

Keywords: Sentiment Analysis, NLP, Neural Networks, Transformers, Sarcasm Detection, Emotion Detection, Web Scraping.

1. INTRODUCTION

In the modern age, where digital trade is booming, ecommerce websites like Amazon continue to tailor consumer behavior by harnessing customer reviews and feedback, making it a significant asset to comprehend customer behavior. They transform how consumers express their opinions, how they communicate and share information online. Millions of reviews are generated every day, these platforms hence provide a warehouse of vast crowd sourced data, with abundant prospects for analysis and insights into consumer insights and behavior [3][4]. Analyzing these reviews is also crucial to businesses to enhance and improve their product, thereby increasing customer satisfaction. Analyzing and comprehending user generated feedback by capturing the emotions, sentiments, sarcasm and context ingrained in these reviews is essential for businesses to keep track of brand perceptions, maintain and improve customer satisfaction, and identify a rising market trends [1][5][7]. This research focuses on customer reviews, keeping in mind how customers articulate and express themselves through positive, negative and neutral reviews. Blending the borders of the conventional systems, this study involves emotion and sarcasm detection to achieve a richer and improved understanding of customer feedback [7][9]. Whether it is identifying emotions such as frustration, anger or disappointment, this multidimensional approach enables companies and consumers to infer significant conclusions [5]. Nevertheless, analyzing these reviews is fundamentally complex due to their brief, abrupt, informal and context-oriented format, highly dependent on the human language [2][6]. They are infused with backhanded comments and underlying emotions, complicating the process. By addressing these gaps, the research aims to provide a deeper sentiment analysis, by employing advanced Natural Language Processing and Machine Learning techniques.

2. LITERATURE SURVEY

Several studies have focused on developing a sentiment analysis model with improving accuracies using different technologies like transformers, NLP etc.

In [1], the researchers used lexicon-based and statistical approaches to identify idiomatic expressions. They also analyzed the polarity using WordNet and other such resources. They combined a lexicon of 580 sentiment bearing idioms through crowdsourcing, ensuring high inter-annotator agreement. They evaluated their model against two existing models using a corpus of sentences containing idioms, annotated with emotions. Though the accuracy was not explicitly reported, the study claimed to improve the sentiment classification when idioms are correctly classified. The F1 score improved by 20 and 15% in two separate experiments, with notable improvements in the classification of positive sentiments, where recall increased by 45% without compromising precision.

Lufti Budi Ilmawan et al. (2024) [2] focused on improving sentiment analysis by handling negations. They emphasized on the usage of syntactic parsing and rule-based handling of negations. This approach aimed to address the challenges posed by negations, which can invert the sentiment of a statement. They also focused on detecting the scope so as to improve the classification. The study showed a 5-10% improvement over the models without handling negations. Santwana Sagnika et al. (2020) [3] focused on multilingual embedding’s (e.g. mBERT) and transfer learning to enable sentiment analysis across various languages. They achieved an accuracy of about 83% across major languages, and slightly lower for low-resource languages. In [4], the researchers applied a hybrid approach, combining machine learning, NLP and rule-based filters on Twitter and Facebook datasets. The study focused on handling noisy and informal language existing in social media content. This study demonstrated an accuracy of approximately 70-75% over traditional systems, on noisy data.

Dinkar Kumar et al. (2024) [5] applied supervised learning (Random Forest and SVM) on social media comments, and combined it with preprocessing pipelines for real-time analysis. They reached an accuracy of about 88.9% with ensemble techniques, on cleaned and preprocessed datasets. In [6], the researchers integrated both textual and visual components from social media applications. They compared various classifiers like SVM, Decision Tree, Random Forest, Logistic Regression, Multinomial Naive Bayes, and Gradient Boosting. They used Flickr30k dataset which contains 31,784 images with captions for training. They achieved an accuracy of 82%.

The authors of [7] detailed their findings by demonstrating how an ensemble classifier, combining CNN, LSTM and rule-based sentiment detects sarcasm in Twitter data. They achieved an accuracy of approximately 84% for detecting sarcasm in online texts. In [8], the authors proposed a transformer-based ST-GCN model; employing TF-IDF and PMI based graph construction, BERT embeddings and spatial-temporal graph learning. They achieved an accuracy of 91.7% across multiple public statement datasets.

Gagan Sharma (2023) [9], stressed on the usage of late fusion by combining the output logits of several transformers (BERT, mBERT, XML-R, Hi-BERT), followed by neural fusion layers. They reported F1 scores of 82.9% for English-Hindi, 84.1% for English-Spanish. In [10], the researchers employed a shared bidirectional LSTM network to capture contextual information from text inputs. It achieved an F1 score of 94%, outperforming existing methods by a margin of 3%.

3. GAPS IDENTIFIED

Although there is a significant progress in sentiment analysis through the use of transformer-based models and ensemble learning, various gaps remain unresolved, especially for e-commerce applications. 1. The baseline models tackle sentiment polarity, sarcasm and emotion as separate tasks, thereby, lack an integrated model to capture their interdependencies [7][8][9]. In spite of the fact that sarcasm detection has been employed in classifying social media texts, it remains marginalized in product feedback, where sarcasm is prevalent [6][7]. Emotion detection is often disregarded as several system just use labels like positive, negative or neutral [1][4]. This makes the outcomes less detailed and not as useful. Moreover, while several systems achieve high accuracy, they lack visual interpretability, thus making them unsuitable for users. Lastly, many of these baseline models are trained on twitter data, thus they struggle with the structured review data like amazon or any other ecommerce applications [5][8][9]. These gaps directly form, the base of our project, which aims to develop a sentiment analysis model tailored for product review platforms.

4. PROPOSED METHODOLOGY

The approach is grounded in conventional sentiment analysis complemented with more sophisticated Natural Language Processing (NLP) and Machine Learning (ML) for the extraction of sentiment, emotion, and sarcasm from Amazon review comments. There are 4 primary steps for this approach: data collection, data preprocessing and feature extraction, detection of sentiment and sarcasm, and visualization. These processes are significant to identify the accuracy and comprehensiveness of the sentiment analysis process.

4.1 Datasets Used

Three publicly available datasets were used – one for training the Sentiment Analysis model, and the other two for training the Sarcasm Detection Model.

Amazon Reviews Dataset- Contains over 500,000 food product reviews, each with a review text, rating (1-5) and meta data such as reviewer ID and date. Star ratings were mapped to the sentiment classes as:

Twitter Sarcasm Corpus- It was used for training the Sarcasm Detection model. It is available on Kaggle and contains short, punctuated and exaggerated sentiment which is useful for training the said model. It contains a number of tweets labelled as “Sarcastic” or “Not Sarcastic.”

Sarcasm Headlines Dataset- This dataset is also available on Kaggle and is used to train the Sarcasm Detection Model. It was used to supplement the training of the model to improve generalization. It consists of 30,000+ headlines from The Onion (sarcastic) and HuffPost (non-sarcastic), each labeled accordingly.

4.2 Data collection

The first step is to collect data from the Amazon website by scraping the reviews of a product. Web scraping extracts text data from reviews and metadata including star rating, date and time stamps, and reviewer. Automatic web scraping of data from other product pages are performed by automated web scraping packages like BeautifulSoup, Scrapy, or Selenium. Data obtained is structured (rating) and unstructured (free-text) data, both being the foundation of sentiment analysis.

4.3 Data Preprocessing and Feature Extraction

Raw data contains special characters, symbols, and contains slangs that impact model performance. So it should be cleaned using various text cleaning approaches which involves some of the following steps:

Step-1: Noise Removal: Special character elimination, HTML tag elimination, special space elimination, and stop words elimination.

Step-2: Tokenization: Word or phrase in text broken up for processing.

Step-3: Lemmatization & Stemming: Shortened to root word with an attempt at making it symmetrical (i.e., “running” into “run”).

Step-4: Normalization: Misspelling, abbreviation, and slang opened into regular words. Feature selection is applied to select the features from the data which are used to feed the machine learning algorithms for better performance. Sentiment Score is extracted, in terms of sentiment score in numerical extraction with VADER polarity score.

4.4 Model Architecture and Training

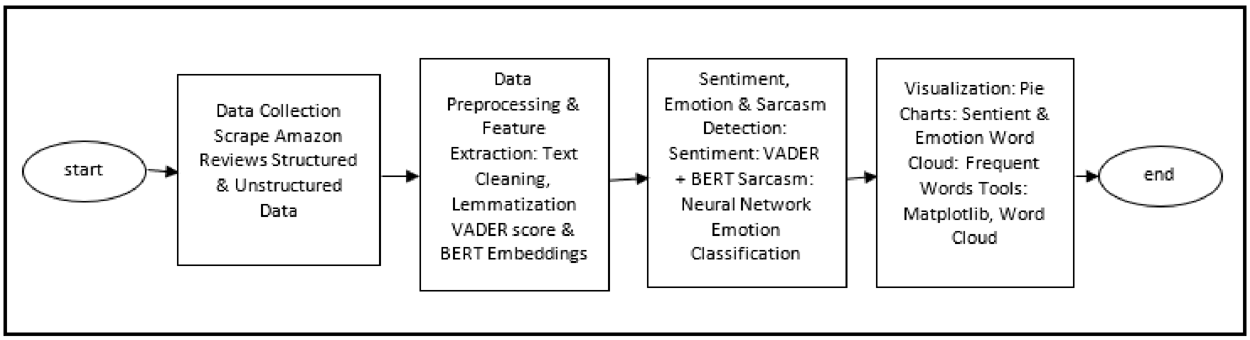

Figure 1: Architecture Diagram

The proposed system consist two main components- a Sentiment analysis model and a sarcasm detection model. These models were trained independently and integrated into a pipeline, to enable a better understanding of customer feedback, especially in e-commerce websites.

4.4.1 Sentiment Analysis Model

An ensemble-based sentiment analysis model was developed using both lexicon-based and transformer-based techniques, so as to classify the reviews into negative, positive, or neutral. First, VADER (Valence Aware Dictionary for Sentiment Reasoning) was used to generate sentiment scores. VADER is a rule-based tool that computes polarity scores– positive, negative, neutral from a compound score– from the input text.

Algorithm 1 VADER Input: Amazon Reviews T
Polarity scores (Positive, Negative, Neutral, and Compound)

1. Tokenize the Text: Break the input text into words and symbols
2. Assign Sentiment Score: For each token
   a. Look up its sentiment score from the VADER lexicon.
   b. If the word is not in the lexicon, score = 0.
3. Adjust Scores Using Rules: For each token with a non-zero score:
   a. Boosters: Words like very, extremely increase the score by ~30%.
   b. Negations: Words like not, never flip the sentiment (~74%).
   c. Capitalization: ALL CAPS words are given more weight.
   d. Punctuation:
      i. Exclamation marks (!) boost intensity.
      ii. Repeated punctuation (e.g., “!!!”) adds more emphasis.
4. Sum and classify Scores
   a. Positive: Sum of all positive scores.
   b. Negative: Sum of all negative scores (as absolute values).
   c. Neutral: Number of words with score = 0.

5. Compute Compound Score: Use this normalization formula:

$$\text{Compound} = \frac{\Sigma scores}{\sqrt{(\Sigma scores)^{2} + 15}}$$

Finally the output value will be in between −1 (more Negative) and +1 (more Positive).

4.4.2 Sarcasm Detection Model

In light of the limitations of the traditional sentiment models in detecting sarcasm, a deep learning model was trained specifically for sarcasm detection. This model was built using a Recurrent Neural Network (RNN) architecture. The training data, as mentioned before, is composed of two datasets- the Twitter Sarcasm Corpus, and The Sarcasm Headlines Dataset.

Preprocessing of this data retained stop-words, punctuation and case sensitivity, to preserve the structure of sarcasm. The texts were tokenized and padded to a fixed length. The RNN architecture had an embedding layer and a standard RNN layer that processed the input sequence in a forward manner. Then came the dropout layer, which was used to mitigate overfitting. Lastly, a dense layer with sigmoid activation was added for binary classification (sarcastic/not sarcastic). The model was trained using binary cross entropy as the loss function and the Adam optimizer. Training was conducted over a limited number of epochs with early stopping based on validation loss. The sarcasm detection model achieved an accuracy and F1 score of 94%, indicating strong generalization performance across several sarcastic expressions.

4.4.3 Model Integration and Inference Pipeline

After training the models individually, they were integrated into a unified inference pipeline and tested within a locally hosted web application using Streamlit. This interface was designed to simulate the user experience of real-time review analysis while maintaining all back-end processing offline.

In the pipeline, user input is first passed to the sarcasm detection model to flag any potentially misleading content. If sarcasm is detected, the sentiment prediction is marked accordingly. The same input is then processed by the sentiment analysis ensemble model, which combines features from VADER and RoBERTa to produce a final sentiment classification. The web interface built with Streamlit enables users to enter review text and immediately visualize the results. Output includes the detected sentiment, sarcasm label, and supporting visuals such as pie charts, bar graphs, and word clouds. These components provide a user-friendly way to interpret complex model outputs.

4.5 Data Visualization and Interpretation

It involves creating interactive dashboards and visualizing the processed data for the user. Sentiment Analysis and Emotion distribution are represented using pie charts, easing the interpretation of the result. A word cloud of all the reviews is displayed, which contains the most used words in the reviews. These are implemented using MatPlotlib and Word Cloud libraries in Python. These help the user to gain better insights into the product, to understand customer sentiment and to detect potential issues.

5. RESULT

The considered sentiment analysis and sarcasm detection system was evaluated for testing using about 500 samples of Amazon product reviews. The main aim was to assess the accuracy, reliability, and interpretability of the models put into use for sentiment classification and sarcasm detection. The sentiment analysis model, employing hybrid features from VADER-a rule-based lexicon and RoBERTa-a deep learning transformer received an overall accuracy of 89%. A look into isolated metrics would reveal RoBERTa was better than VADER due to it being able to capture contextual cues in reviews that were often ambiguous or nuanced. Yet the ensemble has better generalization ability and stability irrespective of the diversity in the review types, and this provides evidence that ensemble learning aids enhancing sentiment prediction by exploiting the advantages of both shallow and deep learning.

The sarcasm detection model, on the other hand, with an architecture based on RNN, had its test accuracy pegged at 94% and an F1-score also of 94%, providing proof for its efficiency in identifying language loaded with sarcasm that generally exists in user feedback. The retention of punctuation, stop words, and case in the text during preprocessing greatly assisted the detection of sarcastic expressions. The visualization layer enables businesses to quickly identify areas of customer satisfaction or concern and adjust product or service strategies accordingly.

Fig.2: Amazon product url as input

Fig.3: Sarcasm

Fig.4: Sentiment and Emotion Distribution

Furthermore, the system supports product-level sentiment comparison, allowing users to analyze and contrast reviews across multiple products. By aggregating sentiment and sarcasm insights for each product, comparative visual reports help stakeholders make data-driven decisions when evaluating product performance and customer satisfaction. Weaving interactive pie charts, bar graphs, and word clouds into the interface is something that can impart substantial intuitive understanding of the results. These graphical objects display the sentiment-class distribution, identify the most frequently occurring terms, and assist in further probing the emotional tone of the reviews. From the perspective of the business, the visualization sheet promptly discloses satisfactory or worrisome factors of their customers to whom they may be able to advance their product or service strategy accordingly. Another key feature of the system is product-level sentiment comparison, which lets users compare reviews for different products. While the system generates the same visualizations for each product as it does in the single product analysis (sentiment, emotion, and sarcasm charts), only a summary table is shown here to avoid repetition. This table compares the number of positive, negative, and sarcastic reviews for each product. A product with more positive and fewer negative or sarcastic reviews is considered to have better overall perception. For demonstration, two similar products were chosen to evaluate how sentiment and sarcasm detection can support real-world decision-making.

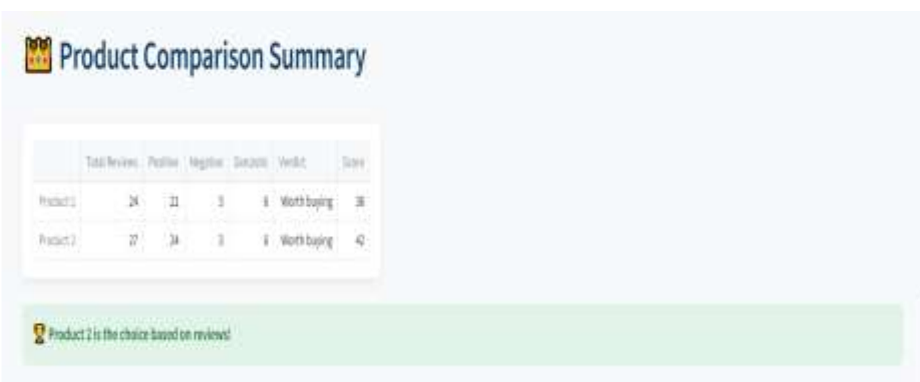

Fig.5: Product Comparison

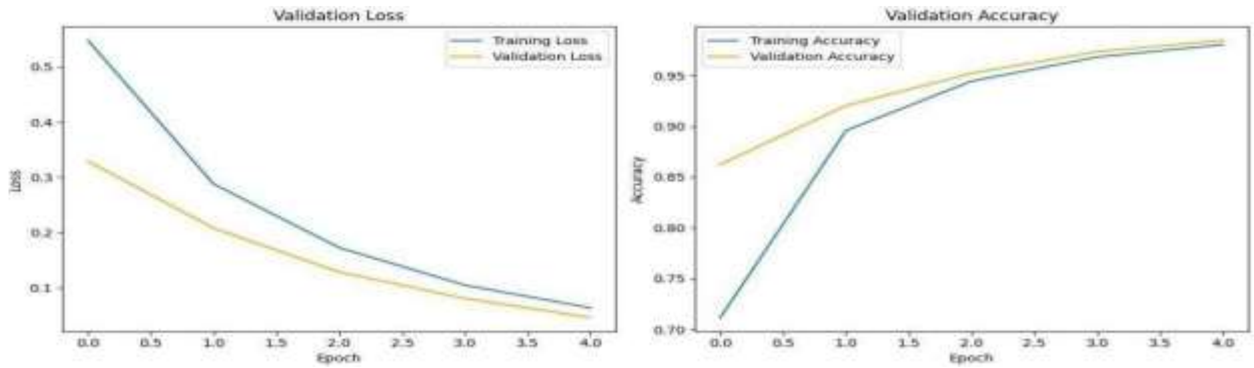

Fig.6: Validation Loss and Accuracy for Sarcasm Detection

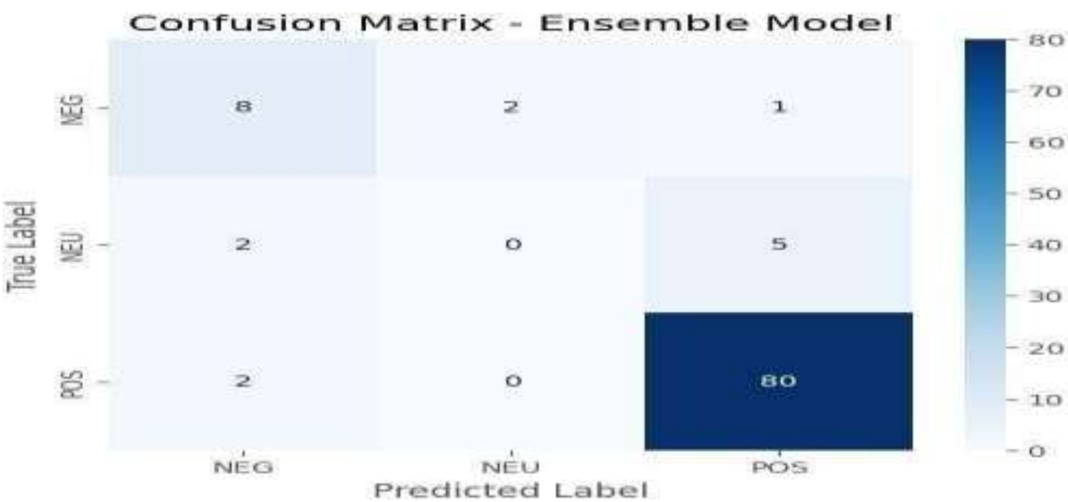

Fig.7: Confusion Matrix for Sentiment Analysis Model

6. COMPARATIVE ANALYSIS:

Criteria	Dinkar Kumar et al. (2023) IJRPR [5]	Sunusi Kabir Alaramma (2023) Sarcasm Detection Focused [7]	Our Proposed Model
Focus Area	General Sentiment Analysis of Social Media	Sarcasm Detection in Social Media	Sentiment, Emotion & Sarcasm Detection in Amazon Reviews
Dataset Used	Twitter data (tweets, hashtags, mentions)	Twitter and Reddit comments	Scraped Amazon product reviews using product URL
Model Type	Machine Learning (SVM, Naive Bayes)	Hybrid Lexicon + Deep Learning (LSTM)	Transformer (BERT) for sentiment & emotion BiLSTM for sarcasm
Preprocessing Steps	Tokenization, Stop word removal, Lemmatization	Slang normalization, POS tagging, emoji handling	Light cleaning; preserves sarcasm cues (stop words, punctuation, emojis)

Sarcasm Handling	Not addressed	Core focus; uses LSTM with emoji patterns	Dedicated sarcasm detection module with BiLSTM and dropout
Multilingual Support	English only	English only	Multilingual input support planned
Accuracy Achieved	82% (Sentiment)	85% (Sarcasm)	88% (Sentiment) 94% (Sarcasm)
Evaluation Metrics	Accuracy, Precision, Recall	Precision, Recall, F1Score	Accuracy, F1-Score, ROCAUC, Confusion Matrix
Visualization Tools	Basic (bar charts, line plots)	None	Interactive pie charts, Sentiment–Emotion Comparison, word clouds
Use Cases	Brand Monitoring, Public Opinion Mining	Humor/Sarcasm interpretation in social media	Product Feedback Analysis,
Key Limitations	No sarcasm detection, poor emoji understanding	Domain-limited (social media only)	Resource-heavy (BERT), limited to product reviews currently

7. CONCLUSION

This paper seeks to integrate state-of-the-art sentiment analysis techniques, including emotion and sarcasm detection, to best address the multivalence of Amazon review comment text. By strict application of highly sophisticated text embedding and classification technology, this study improves sentiment class performance and detects latent emotional signals such as elation, frustration, and sarcasm that traditional methods omit. The study uncovers strong correlations among consumer attitude trends and important measures of product success, such as sales and ratings, that are directly actionable. In conclusion, the process offers the organizations a tool to understand the behavior of the customer, which in turn allows informed decision making and competitiveness in the market.

8. REFERENCES

1. Lowri Williams et al. – The Role of Idioms in Sentiment Analysis – 2015, Expert Systems with Application.
2. Lutfi Budi Ilmawan et al. – Understanding Negations in Sentiment Analysis – 2024, ResearchGate
3. Santwana Sagnika et al. – Multilingual Sentiment Analysis – 2020, ResearchGate.
4. Shanthi, Dr. D., G. Ashok, Chitrika Biswal, Sangem Udharika, Sri Varshini, and Gopireddi Sindhu. 2025. “Ai-Driven Adaptive It Training: A Personalized Learning Framework For Enhanced Knowledge Retention And Engagement”. Metallurgical and Materials Engineering, May, 136-45. https://metall-mater-eng.com/index.php/home/article/view/1567.
5. Federico Neri, Carlo Aliprandi, Federico Capeci, Montserrat Cuadros, Tomas By – Sentiment Analysis on Social Media – 2012, IEEE
6. Dinkar Kumar et al. – Sentiment Analysis of Social Media Presence – 2023, International Journal of Research Publication and Reviews
7. Dinkar Kumar, Kundan Kumar, Kritim Prasad Kafle, Vishal Kumar, Jivesh Kumar, A. Periyasamy – Sentiment Analysis of Social Media Presence – 2024, International Journal of Research Publication and Reviews
8. Sunusi Kabir Alaramma – Sentiment Analysis of Sarcasm Detection in Social Media – 2023
9. Barakat AlBadan – Transformer-Based Graph Convolutional Network for Sentiment Analysis – 2022
10. Gagan Sharma – Late Fusion of Transformers for Sentiment Analysis of Code-Switched Data – 2023
11. Yik Yang Tan, Chee-Onn Chow, Jeevan Kanesan, Joon Huang Chuah, YongLiang Lim – Sentiment Analysis and Sarcasm Detection using Deep Multi-Task Learning – 2023, Wireless Personal Communications, Springer. [DOI: 10.1007/s11277-023-10235-4]
12. Mayur Wankhade, Annavarapu Chandra Sekhara Rao, Chaitanya Kulkarni – A Survey on Sentiment Analysis

Methods, Applications, and Challenges – 2022, Springer.

13. D Shanthi, "Smart Water Bottle with Smart Technology", Handbook of Artificial Intelligence, Benthem Science Publishers, Pg. no: 204-219, 2023
14. D Shanthi , Smart Healthcare for Pregnant Women in Rural Areas, Medical Imaging and Health Informatics, Wiley Publishers,ch-17, pg.no:317-334, 2022
15. D. Shanthi, R. K. Mohanty and G. Narsimha, "Application of machine learning reliability data sets", Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS), pp. 1472-1474, 2018.
16. D Shanthi, N Swapna, Ajmeera Kiran and A Anoosha, "Ensemble Approach Of GPACOTPSOAnd SNN For Predicting Software Reliability", International Journal Of Engineering Systems Modelling And Simulation, 2022.
17. D. Shanthi, "Ensemble Approach of ACOT and PSO for Predicting Software Reliability", 2021 Sixth International Conference on Image Information Processing (ICIIP), pp. 202-207, 2021.
18. D Shanthi, CH Sankeerthana and R Usha Rani, "Spiking Neural Networks for Predicting Software Reliability", ICICNIS 2020, January 2021, [online] Available: https://ssrn.com/abstract=3769088.
19. Shanthi, D. (2023). Smart Water Bottle with Smart Technology. In Handbook of Artificial Intelligence (pp. 204-219). Bentham Science Publishers.
20. D. Shanthi, P. Kuncha, M. S. M. Dhar, A. Jamshed, H. Pallathadka and A. L. K. J E, "The Blue Brain Technology using Machine Learning," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1370-1375, doi: 10.1109/ICCES51350.2021.9489075.
21. Shanthi, D., Aryan, S. R., Harshitha, K., & Malgireddy, S. (2023, December). Smart Helmet. In International Conference on Advances in Computational Intelligence (pp. 1-17). Cham: Springer Nature Switzerland.