Journal of Theoretical and Applied Information Technology
15th November 2024. Vol.102. No. 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ENHANCING USER SAFETY: AI-DRIVEN POPUP ALERTS FOR SUSPICIOUS CONTENT USING DEEP MULTI-MODAL ANOMALY FUSION NETWORK

B SHANTHINI[1], Dr. N SUBALAKSHMI[2]

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Tamil Nadu,India

[2]Assistant Professor, Department of Computer and Information Science, Annamalai University,TamilNadu, India

E-mail: [1]shanthinib.66@gmail.com , [2]subhaabi12@gmail.com

ABSTRACT

This research introduces a novel approach to enhancing user safety in online environments through AI-driven popup alerts for detecting suspicious content. By using a Deep Multi-Modal Anomaly Fusion Network, we integrate features extracted from text, emojis, images, and videos to construct a comprehensive representation of product reviews across diverse modalities. In particular, we employ the innovative AdaptiMatrixFactorizer for feature extraction, which dynamically adjusts its factorization process to capture evolving data patterns. The proposed fusion network architecture seamlessly integrates features from different modalities, utilizing concatenation, element-wise addition, and attention mechanisms to facilitate effective multi-modal anomaly detection using Deep Multi-Modal Anomaly Fusion Network (DMMAFN). Furthermore, our approach introduces a dynamic threshold adjustment mechanism within the fusion network to adaptively regulate anomaly detection sensitivity based on real-time changes in data patterns and distributions. This adaptive thresholding strategy considers three critical parameters: sentiment analysis, repetitive reviews, and spatio-temporal analysis. Upon detection of suspicious content exceeding the dynamically adjusted threshold, a popup alert is generated to notify users, fostering a safer online environment. Extensive experimentation and evaluation on anomaly-labeled datasets demonstrate the efficacy and reliability of our approach in accurately detecting and alerting users to potential risks in product reviews across various online platforms. This research contributes to advancing user safety in online environments by providing a proactive and intelligent solution for identifying and addressing suspicious content, while also introducing the novel AdaptiMatrixFactorizer for dynamic feature extraction.

Keywords: *AI-driven Popup Alerts, Suspicious Content Detection, Deep Multi-Modal Anomaly Fusion Network, AdaptiMatrixFactorizer, Sentiment Analysis.*

1. INTRODUCTION

In the present era of technology, the internet has become an essential component of our everyday existence, providing us with ease, connectedness, and the ability to access a wide range of information and services. However, with the proliferation of online platforms and the exponential growth of user- generated content, ensuring user safety in online environments has become an increasingly complex challenge. [1] One of the key concerns is the presence of suspicious or harmful content, such as fake reviews, spam, misinformation, and inappropriate material, which can negatively impact user experience and trust in online platforms [2]. In order to deal with these difficulties, this study presents an innovative method leveraging Artificial Intelligence (AI) to enhance user safety through proactive detection of suspicious content [3]. Our proposed solution utilizes AI-driven popup alerts to promptly notify users when encountering potentially harmful content, thereby empowering them to make informed decisions and navigate online platforms with confidence. Central to our approach is the use of a Deep Multi-Modal Anomaly Fusion Network, a sophisticated AI architecture capable of integrating features extracted from various modalities, including text [4], emojis [5], images [6], and videos [7]. By constructing a comprehensive representation of product reviews across diverse modalities, our system can effectively identify

anomalous patterns indicative of suspicious content. A key innovation in our methodology is the utilization of the AdaptiMatrixFactorizer for feature extraction. This dynamic feature extraction algorithm [8] dynamically adjusts its factorization process to capture evolving data patterns, ensuring that our system remains adaptive and responsive to changes in the online landscape. The AdaptiMatrixFactorizer enables us to extract rich and meaningful features from product reviews, facilitating accurate anomaly detection and popup alert generation. Moreover, our proposed fusion network architecture seamlessly integrates features from different modalities using element-wise addition [9].

This integration enables effective multi- modal anomaly detection, allowing our system to capture complex relationships and patterns that may span across multiple modalities. Furthermore, to enhance the robustness and adaptability of our system, we introduce a dynamic threshold adjustment mechanism within the fusion network. This mechanism dynamically regulates anomaly detection sensitivity based on real-time changes in data patterns and distributions, considering three critical parameters: sentiment analysis, repetitive reviews, and spatio-temporal analysis. By dynamically adjusting the threshold [10], our system can effectively distinguish between normal and suspicious content, minimizing false positives and false negatives. In summary, this research contributes to advancing user safety in online environments by providing a proactive and intelligent solution for identifying and addressing suspicious content. Through extensive experimentation and evaluation on anomaly-labeled datasets, we demonstrate the efficacy and reliability of our approach in accurately detecting and alerting users to potential risks in product reviews across various online platforms.

Additionally, by introducing the novel AdaptiMatrixFactorizer for dynamic feature extraction, we pave the way for future advancements in AI-driven safety solutions for online platforms. The paper is organized as follows: Section 1 offers an introductory overview of the research topic and presents a detailed summary of the suggested solution. Section 2 presents a review of related work in the field of user safety and anomaly detection in online environments. Section 3 describes the methodology, including the Deep Multi-Modal Anomaly Fusion Network architecture and the AdaptiMatrixFactorizer for feature extraction. Finally, Section 4 presents experimental results and evaluation metrics, followed by a discussion of the findings and conclusions in Section 5.

2. RELATED WORKS

Manuscripts must be in English (all figures and The research implements NLP and LSTM to develop a consumer review summarization model, featuring a hybrid sentiment analysis approach with pre-processing techniques such as tokenization, stop- word removal, and stemming [11]. Feature extraction involves constructing a hybrid feature vector with review-related and aspect-related features. This study introduces a framework for product recommendation by combining Long short-term memory (LSTM) and collaborative filtering (CF), utilizing an LSTM-based model for SA trained on large-scale product review datasets [12]. Deep learning-based sentiment analysis on Twitter messages is explored, using three-word embeddings schemes and a convolutional neural network (CNN) for feature extraction [13]. The study preprocesses the Twitter messages by tokenization, lowercasing, and removing stop words. A hybrid LSTM-CNN model is proposed for sentiment analysis in Assamese, utilizing Keras word embeddings for vectorization. The model architecture combines LSTM and CNN layers followed by dense layers for classification [14].

The paper investigates sentiment analysis in product reviews using transformer models on the Amazon dataset, comparing various AI-based techniques including BERT [15]. Preprocessing involves tokenization and padding, and the models are fine-tuned on the labeled dataset. This paper reviews recent studies on sentiment analysis using deep learning, comparing word embedding on various datasets including movie reviews and product reviews. Different deep learning architectures are discussed [16]. Evaluation is conducted using standard performance metrics and comparisons are made with traditional machine learning classifiers. A novel context-aware, deep-learning-driven approach for Persian sentiment analysis is presented, utilizing CNN and LSTM models trained on a manually annotated dataset of Persian movie reviews. Preprocessing involves tokenization and word embedding, followed by feature extraction using CNN and LSTM layers [17]. A deep learning-based technique, BiLSTM, is proposed for classifying Bengali restaurant reviews into positive and negative polarities, achieving high accuracy [18]. Preprocessing steps include tokenization, stop-word removal, and word embedding. The methodology involves collecting

and preprocessing customer reviews from multiple sources, training sentiment analysis models using deep learning techniques such as CNN and LSTM, and evaluating the models on metrics like accuracy and precision [19]. A framework for sentiment classification using Recurrent Neural Network (RNN) and LSTM is proposed, evaluated on various datasets with LSTM showing the best performance [20]. The methodology involves data preprocessing including tokenization and padding, model training using RNN and LSTM architectures. The paper explores sentiment analysis techniques, including deep learning models like BERT, RoBERTa, and DistilBERT, on Twitter's US airlines sentiment dataset [21]. Preprocessing involves tokenization, padding, and fine-tuning the transformer models on the labeled dataset. Sentiment polarity detection is performed using different techniques, including LSTM, on Amazon and Yelp customer reviews, with LSTM achieving high accuracy [22].

The methodology involves preprocessing steps such as tokenization and word embedding, model training using LSTM architecture. Various deep learning approaches, including CNN and RNN, are discussed for Twitter sentiment analysis, with models trained and evaluated for accuracy. The methodology involves data preprocessing including tokenization and embedding, model training using CNN and RNN architectures [23]. The paper introduces a CNN-based classifier for sentiment analysis, comparing its performance with other machine learning techniques on manually annotated datasets from IMDB and Amazon [24]. Preprocessing involves tokenization and word embedding, followed by model training using CNN architecture. The methodology involves preprocessing steps such as tokenization and word embedding, model training using LSTM-RNN architecture with attention layers, and evaluation using metrics like accuracy, precision, recall, and F1 score [25].

3. PROPOSED MODEL

The proposed methodology aims to enhance user safety in online environments through AI-driven popup alerts for detecting suspicious content and the proposed architecture is shown in fig 1. Leveraging a DMMAFN, features extracted from text, emojis, images, and videos are integrated to create a comprehensive representation of product reviews.

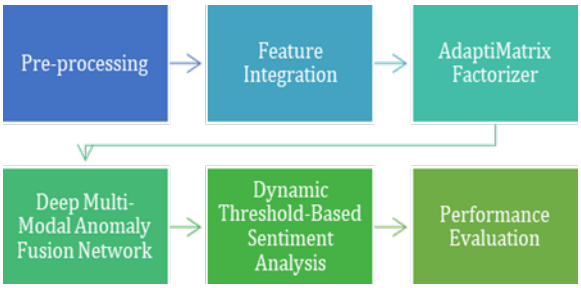

Figure 1: Overall Architecture of Proposed Model

The innovative AdaptiMatrixFactorizer dynamically adjusts its factorization process to capture evolving data patterns effectively. Using concatenation, element-wise addition, and attention mechanisms, the fusion network facilitates multi-modal anomaly detection. A dynamic threshold adjustment mechanism regulates anomaly detection sensitivity based on real-time changes in data patterns, considering parameters such as sentiment analysis, repetitive reviews, and spatio-temporal analysis. Upon detection of suspicious content, popup alerts notify users, fostering a safer online environment.

3.1 Pre-processing

The preprocessing steps on input data, including text, emojis, images, and videos, such as tokenization, removal of special characters, lowercasing, stopword removal, and lemmatization/stemming for text data. For emojis, convert them into textual representations. For images and videos, apply techniques like resizing, normalization. Each step is detailed below: Tokenization splits the input text X into individual tokens.

$$T1, T2, \ldots, Tn : X = T1+T2+\cdots+Tn \quad (1)$$

The feature set is displayed as a matrix, with each row corresponding to a data occurrence and each column representing a feature.

$$\text{Filtered Text} = X - \{\text{special characters}\} \quad (2)$$

Convert all text to lowercase to standardize the text data and avoid duplication of words due to case differences. All tokens are converted to lowercase:

$$\text{Lower cased Text} = \text{lower (Filtered Text)} \quad (3)$$

Stopwords removal involves comparing each token against a predefined list of stopwords and removing them if it matches:

$$\text{Filtered Text} = \text{Lower cased Text} - \{\text{stopwords}\} \quad (4)$$

Emojis are converted into textual representations to ensure consistency in text processing. Each emoji

is replaced with its corresponding textual representation. emoji conversion involves mapping each emoji to its textual representation:

Emoji = Textual Representation (5)

For images and videos to ensure uniformity and compatibility with the model architecture. image and video preprocessing involves operations such as resizing to a standard size and normalization of pixel values:

Preprocessed Image/Video = Transformation (Original Image/Video) (6)

These preprocessing steps are essential for cleaning and standardizing the input data, preparing it for further analysis and modeling. This model helps remove noise, reduce dimensionality, and improve the efficiency and effectiveness of downstream tasks such as feature extraction and model training.

ALGORITHM: PREPROCESSING

Input: Raw data containing text, emojis, images, and videos

Output: Preprocessed data

1. Text Preprocessing:

- Tokenize, remove special characters, lowercase, remove stopwords

2. Emoji Preprocessing:

- Convert emojis to textual representations

3. Image and Video Preprocessing:

- Resize and normalize images and videos

Return preprocessed_text,
preprocessed_emojis,
preprocessed_images,
preprocessed_videos.

The algorithm preprocesses raw data comprising text, emojis, images, and videos for subsequent feature extraction. Initially, the text undergoes tokenization, special character removal, lowercasing, and stopwords removal. Emojis are converted to textual representations. Images and videos are resized and normalized. The preprocessed data, including text, emojis, images, and videos, is then outputted for further analysis and featureextraction.

3.2 Feature Integration

The objective of feature integration is to combine preprocessed features from various modalities (text, emojis, images, and videos) into a unified representation. This unified representation encapsulates the essential information from all modalities and serves as input to the anomaly detection model.

Let, Xemojis, Ximages, and Xvideos denote the preprocessed features extracted from text, emojis, images, and videos, respectively.

The features from different modalities are concatenated to create a single feature matrix Xconcat. This concatenation process preserves the individual features from each modality while combining them into a unified representation.

Xconcat = [Xtext, Xemojis, Ximages, Xvideos] (7)

The feature integration process combines preprocessed features from text, emojis, images, and videos into a unified representation using concatenation. This unified representation is then fed into a Deep Multi-Modal Anomaly Fusion Network, which learns to capture the underlying patterns and relationships between features from different modalities. Anomaly detection is subsequently performed to identify suspicious content in the product reviews based on the integrated feature representation.

3.3 AdaptiMatrixFactorizer

The AdaptiMatrixFactorizer is a novel algorithm used for feature extraction from preprocessed text data. It dynamically adjusts its factorization process to capture evolving data patterns effectively. The AdaptiMatrixFactorizer decomposes the text data matrix X into two low-rank matrices W and H, representing the latent features of the documents and tokens, respectively.

$$X \approx W \times H^T \quad (8)$$

where W is a m×k matrix and H is a n×k matrix, and k is the desired number of latent features.

The AdaptiMatrixFactorizer dynamically adjusts its factorization process to capture evolving data patterns. This adaptive process enables the model to accommodate variations in the data distribution and catch emerging trends. The dynamic adjustment process can be represented by updating the factorization matrices W and H iteratively based on the latest data patterns. After factorization, the latent features extracted by the AdaptiMatrixFactorizer represent the essential characteristics of the text data. These latent features capture semantic and contextual information from the documents and tokens. The latent feature

representation can be obtained by multiplying the factorization matrices:

$$\text{Latent Features} = W \times H^T \quad (9)$$

The latent features extracted by the AdaptiMatrixFactorizer can be integrated with features from other modalities using fusion techniques like emojis [ ], images [ ], videos [ ], creating a comprehensive representation of the product reviews across diverse modalities.

ALGORITHM: ADAPTIMATRIXFACTORIZER

Input: Preprocessed text data matrix X, Desired number of latent features k

Output: Factorization matrices W and H representing latent features of documents and tokens

1. Initialize random matrices W and H with dimensions m x k and n x k, respectively.

2. Set the maximum number of iterations for the factorization process.

3. Repeat until convergence:

a. Compute the reconstruction error E as the Frobenius norm of the difference between X and the matrix product of W and H.

b. Update matrix W: W = W * (X * H') / (W * H * H').

c. Update matrix H: H = H * (W' * X) / (W' * W * H).

d. Compute the reconstruction error at the end of each iteration.

e. If the change in reconstruction error is below a predefined threshold or the maximum number of iterations is reached, stop the process.

4. Return the factorization matrices W and H representing the latent features of documents and tokens.

It then iteratively updates these matrices to minimize the reconstruction error, adjusting the factorization process dynamically. This adjustment mechanism allows the model to capture evolving data patterns effectively. After convergence, the factorization matrices W and H contain the latent features extracted from the text data, capturing semantic and contextual information from the documents and tokens.

3.4 Deep Multi-Modal Anomaly Fusion Network

The Deep Multi-Modal Anomaly Fusion Network is a sophisticated neural network architecture designed to integrate and analyze features from multiple modalities simultaneously. The DMMAFN aims to detect anomalies or suspicious patterns in data that come from diverse sources or modalities. This can be represented as:

$$\text{Anomaly Detection} : \text{DMMAFN}(\text{Multi-Modal Features}) \rightarrow \text{Anomaly Score}$$

The DMMAFN typically consists of multiple layers of neural network units, including convolutional layers, recurrent layers, and fully connected layers. It accepts inputs from multiple modalities and processes them in parallel or through shared layers to extract high-level representations. Let X1,X2,…,Xn denote the input features from different modalities. The forward propagation through the network can be represented as:

$$\text{Output} = Lm(\ldots(L2(L1([X1,X2,\ldots,Xn])))\ldots) \quad (10)$$

The network utilizes fusion techniques to combine features from different modalities effectively. Element-wise addition combines features by adding them element-wise. Fusion layers merge information from various modalities while preserving the unique characteristics of each modality. Fusion techniques combine features from different modalities effectively. Let F(X1,X2,…,Xn) denote the fused representation. fusion mechanisms can be represented as:

$$F(X1,X2,\ldots,Xn) = \text{Element-wise Addition} \quad (11)$$

Element-wise addition combines features by adding them element-wise. However, for concatenation, each modality's feature should have the same dimensionality. Let's denote Xi,j as the j-th feature of modality i, then element-wise addition can be represented as:

$$F(X1 ,X2,\ldots,Xn) = X1,1+ X2,1+ \ldots+ X(n,1),$$

$$X1,2+ X2,2+ \ldots+ X(n,2),$$

$$\vdots$$

$$X(1,m)+X(2,m)+\ldots+X(n,m) \quad (12)$$

Here, m represents the number of features in each modality. This approach emphasizes the relationship between features from different modalities by directly merging them. Mathematically, given features X_1 ,X_2,…,X_n from n modalities, the fused representation F using element-wise addition.

The DMMAFN learns to extract hierarchical features from each modality through the use of convolutional and recurrent layers. These layers capture both low-level and high-level features, allowing the network to understand complex relationships within and across modalities. Let H(X) represent the learned features from a single modality X Feature extraction involves learning hierarchical features through convolutional and recurrent layers:

$$H(X)=Lm(\ldots(L2(L1(X)))\ldots) \quad (13)$$

The final layers of the DMMAFN are responsible for anomaly detection. The modelanalyzes the fused representations from all modalities to identify deviations or anomalies. Anomaly detection mechanisms may involve threshold-based methods, reconstruction error analysis, or deep learning-based approaches. Let A(F) denote the anomaly score. Mathematically, this can be represented as:

$$\text{Anomaly Score} = A(F(X1,X2,\ldots,Xn)) \quad (14)$$

The DMMAFN can adapt to changes in data distribution or the emergence of new anomaly patterns over time. It achieves adaptability through techniques called dynamic threshold adjustment.

ALGORITHM: DEEP MULTI-MODAL ANOMALY FUSION NETWORK

Input: Multi-modal features X1, X2, ..., Xn; Fusion mechanism (e.g., element-wise addition); Anomaly detection mechanism (e.g., threshold-based methods)

Output: Anomaly score indicating the likelihood of suspicious patterns in the input data

1. Initialize the DMMAFN architecture:

• Define layers L1, L2, ..., Lm, including convolutional, recurrent, and fully connected layers.

• Specify the fusion mechanism for combining features from different modalities.

2. Perform forward propagation through the network:

• Concatenate input features from different modalities: [X1, X2, ..., Xn]

• Pass the concatenated features through the layers of the DMMAFN: Lm(...(L2(L1([X1, X2, ..., Xn])))...)

3. Apply fusion mechanism to combine features:

• If using element-wise addition:

• Initialize fused representation F as zeros with the same shape as the input features.

• For each modality i (1 to n):

• Add the features Xi to the fused representation F: F = F + Xi

4. Extract hierarchical features from each modality:

• For each modality X:

• Pass the features through the convolutional and recurrent layers to learn hierarchical representations: H(X) = Lm(...(L2(L1(X)))...)

5. Perform anomaly detection:

• Analyze the fused representation F to identify deviations or anomalies:

• Use the anomaly detection mechanism to calculate the anomaly score A(F) based on the fused features: Anomaly Score = A(F)

6. Return the anomaly score indicating the likelihood of suspicious patterns in the input data.

The algorithm initializes the DMMAFN architecture and performs forward propagation through the network to extract features from different modalities. It then applies a fusion mechanism, such as element-wise addition, to combine features effectively while preserving their unique characteristics. Next, hierarchical features are extracted from each modality using convolutional and recurrent layers. Finally, the model analyses the fused representation to detect anomalies and calculates an anomaly score indicating the likelihood of suspicious patterns in the input data.

3.5 Dynamic Threshold-Based Sentiment Analysis

Journal of Theoretical and Applied Information Technology
15th November 2024. Vol.102. No. 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

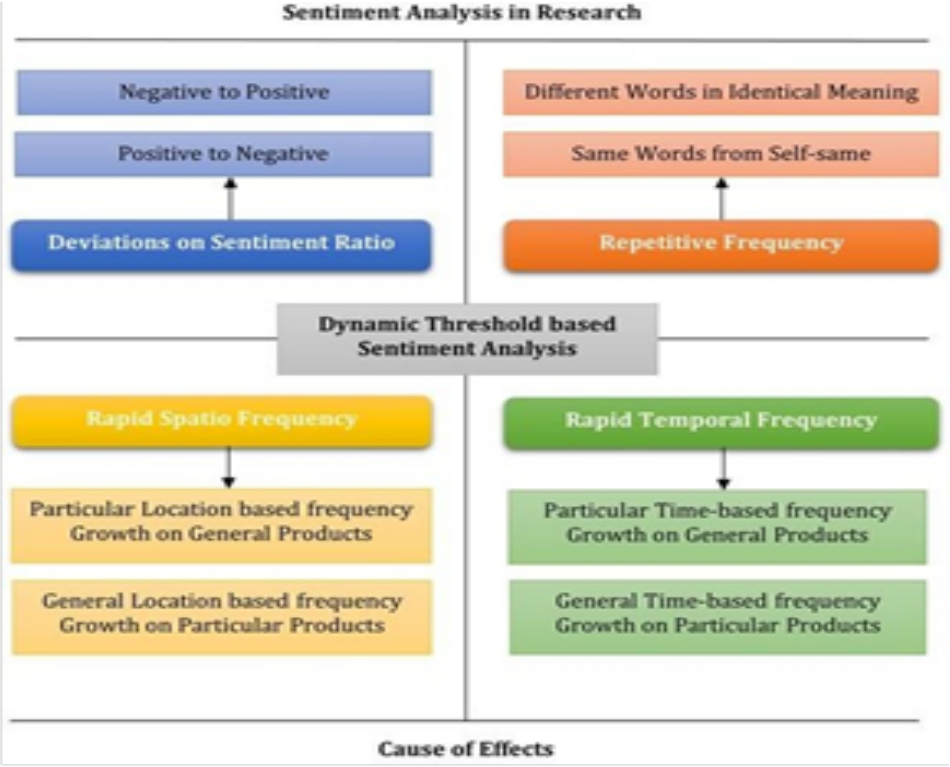

Figure 2: Dynamic Threshold based Sentiment Analysis

The dynamic threshold-based sentiment analysis in our proposed model aims to adaptively regulate anomaly detection sensitivity based on real- time changes in data patterns and distributions as shown in fig 2. This adaptive thresholding strategy considers three critical parameters: sentiment analysis, repetitive reviews, and spatio-temporal analysis l.

i. Sentiment Analysis Thresholding

In the context of detecting anomalies in sentiment expression, we can adjust the sentiment threshold Tsentiment to identify instances where a review inaccurately conveys positivity or negativity. Specifically, we aim to dynamically adjust the threshold based on the standard deviation of sentiment scores (S). Let's consider the scenario where a positive comment is misclassified as negative or vice versa. In such cases, the sentiment score S would deviate significantly from the expected sentiment. We can dynamically adjust the sentiment threshold Tsentiment as follows:

• If the sentiment score S of a review deviates significantly from the mean sentiment score, it indicates a potential anomaly in sentiment expression.

• We adjust Tsentiment by adding a scaled factor of the standard deviation of sentiment scores to the mean sentiment score. Mathematically, the adjusted sentiment threshold Tsentiment is calculated as:

$$Tsentiment = mean(S) + k \times std(S) \quad (15)$$

where k is a scaling factor that determines the sensitivity of the threshold adjustment.

By dynamically adjusting Tsentiment based on the standard deviation of sentiment scores, we can effectively identify instances where the sentiment expressed in a review deviates significantly from the norm, indicating potential anomalies in sentiment expression.

ii. Repetitive Reviews Thresholding

To detect repetitive reviews, the system monitors the frequency of reviews from the same user or containing duplicated content. Let R represent the frequency ratio of repetitive reviews. The threshold Trepetitive is dynamically adjusted based on the standard deviation of the frequency ratio:

$$Trepetitive = mean(R) + k \times std(R) \quad (16)$$

iii. Spatio-Temporal Analysis Thresholding

For spatio-temporal analysis, let F represent the frequency ratio of reviews from specific user

locations or timestamps. The threshold Tspatio−temporal is dynamically adjusted based on the standard deviation of the frequency ratio:

$$Tspatio - temporal = mean(F) + k \times std(F) \quad (17)$$

By integrating these dynamic threshold adjustments, the system ensures prompt and accurate detection of suspicious content, thereby enhancing user safety in online environments. The thresholds are continuously updated based on the evolving data patterns, allowing the system to adapt to changing conditions and effectively detect anomalies in sentiment expression, repetitive reviews, and spatio- temporal behavior.

ALGORITHM: DYNAMIC THRESHOLD-BASED ANOMALY DETECTION

Input: Sentiment scores (S) for each review; Frequency ratio (R) of repetitive reviews; Frequency ratio (F) of reviews from specific user locations or timestamps; Scaling factor (k) for threshold adjustment
Output: Adjusted thresholds for sentiment analysis (Tsentiment), repetitive reviews (Trepetitive), andspatio-temporal analysis (Tspatio-temporal)

1. Sentiment Analysis Thresholding:
   - Calculate the mean sentiment score: mean_S = mean(S)
   - Calculate the standard deviation
2. Repetitive Reviews Thresholding:
   - Calculate the mean frequency ratio ofrepetitive reviews: mean_R = mean(R)
   - Calculate the standard deviation of the frequency ratio: std_R = std(R)
3. Spatio-Temporal Analysis Thresholding:
   - Calculate the mean frequency ratio of reviews from specific user locations or timestamps: mean_F = mean(F)
   - Calculate the standard deviation of the frequency ratio: std_F = std(F)
4. Return the adjusted thresholds Tsentiment, Trepetitive, and Tspatio-temporal.

This algorithm dynamically adjusts thresholds for sentiment analysis, repetitive reviews, and spatio-temporal analysis based on real-time changes in data patterns and distributions. It calculates the mean and standard deviation of sentiment scores, frequency ratios of repetitive reviews, and frequency ratios of reviews from specific user locations or timestamps. Then, it adjusts the thresholds by adding a scaled factor of the standard deviation to the mean value. The workflow of proposed model is shown in fig 3.

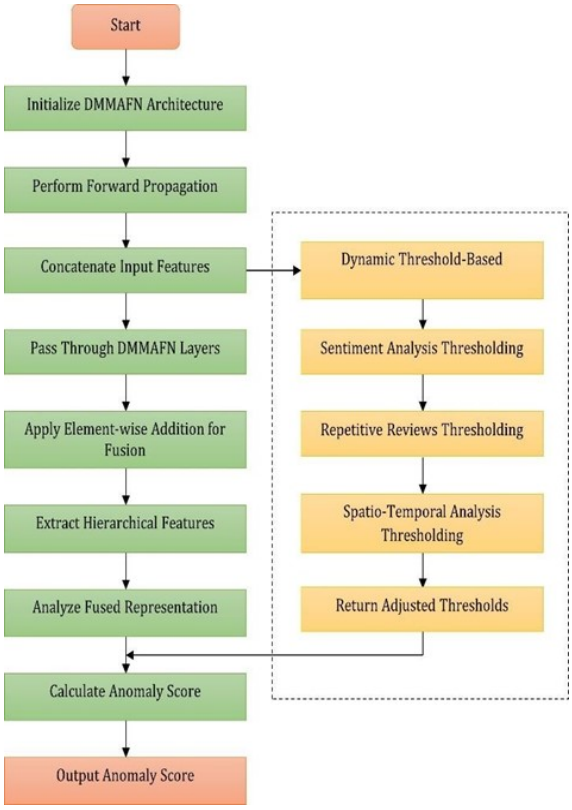

Figure 3: Workflow of Proposed Model

This approach ensures that the thresholds adapt to evolving data patterns, allowing the system to effectively detect anomalies in sentiment expression, repetitive reviews, and spatio-temporal behavior.

4. RESULTS AND DISCUSSIONS

4.1 Dataset Description

Amazon dataset was collected from: https://www.kaggle.com/datasets/lokeshparab/amazo n-products-dataset. The sample product data is categorized into 142 distinct categories and is stored in .csv format, much like the whole dataset. Figure 4 displays the standardization of the top twenty brands in the Amazon dataset. The original dataset from Amazon platform were collected from: https://www.kaggle.com/datasets/daishinkan002/ama zon-mobile-dataset.

Journal of Theoretical and Applied Information Technology
15th November 2024. Vol.102. No. 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

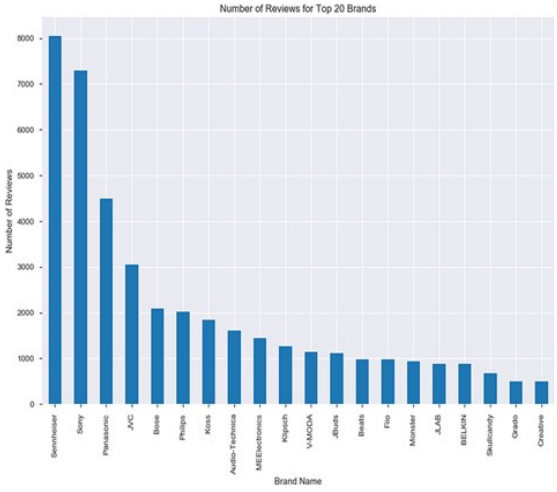

Figure 4: Normalization of Number of Reviews in AmazonDataset

The dataset contains the following columns: Title, Price, Availability Status, Rating, Total Reviews, Reviews, Reviews Rating and Product Description. The frequency distribution based on number of reviews in the Amazon dataset is shown in fig 5 and Table 1 shows the dataset descriptions were collected from Amazon.

Table 1: Overall Comparison of Performance Evaluation

Feature Name	Description
Title	The name or title of the mobile product listed on Amazon.
Price	The price of the mobile product in the respective currency.
Availability Status	The current stock status of the mobile product (e.g., "In Stock," "Out of Stock").
Rating	The overall customer rating of the mobile product, usually on a scale of 1 to 5 stars.
Total Reviews	The total number of customer reviews for the mobile product.
Reviews	A list or collection of individual customer reviews for the mobile product.
Reviews Rating	The individual rating given by customers in their reviews, usually on a scale of 1 to 5 stars.
Product Description	A detailed description of the mobile product, including its features, specifications, and other relevant information.

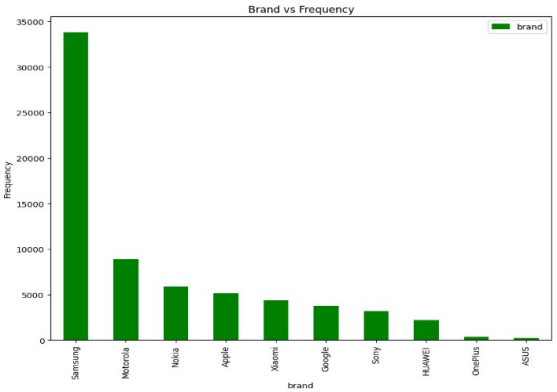

Figure 5: Frequency Distribution in Amazon Dataset

4.2 Experimental Analysis

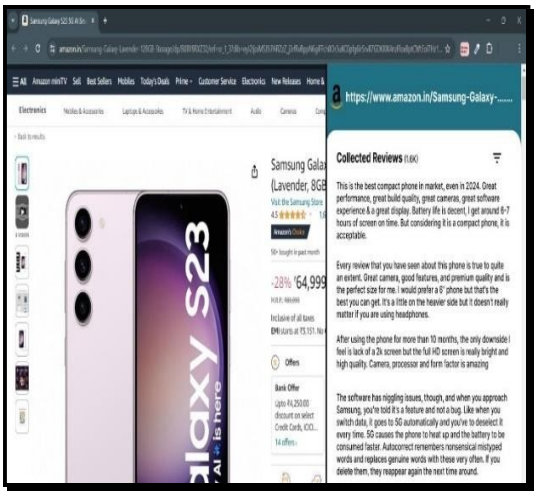

Figure 6: Collected Reviews from Amazon

Figure 6 illustrates the reviews collected from Amazon website. The diagram shows the distribution and variety of reviews, which contains both safe and unsafe feedback. This visualization aids in understanding customer sentiment and the overall reception of products on the platform.

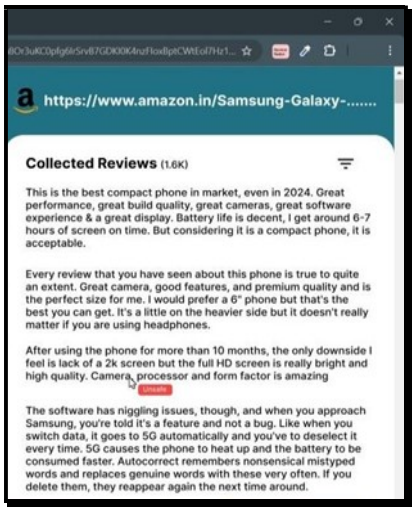

Figure 7: Pop up Alerts for Unsafe Reviews

Figure 7 showcases pop-up alerts generated for unsafe reviews, providing timely notifications to users. In contrast, Figure 8 displays pop-up alerts for safe reviews, ensuring users are informed about trustworthy content. Figure 9 presents an overall summary of anomaly detection processes, offering insights into the effectiveness of the detection system.

Journal of Theoretical and Applied Information Technology
15th November 2024. Vol.102. No. 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

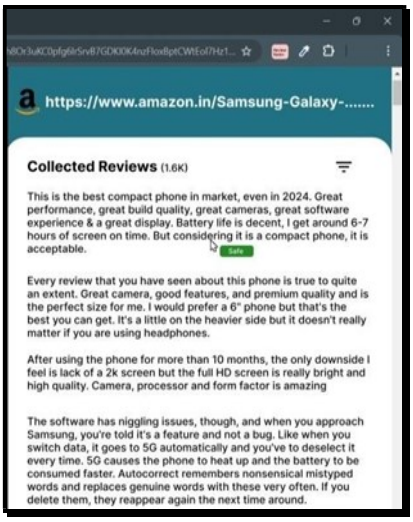

Figure 8: Pop up Alerts for Safe Reviews

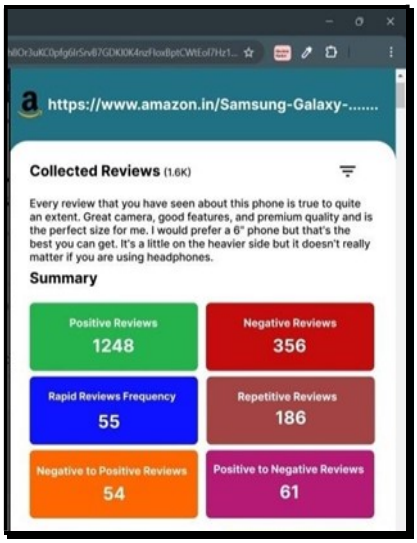

Figure 9: Overall Summary of Anomaly Detection

4.3 Performance Evaluation

Sentiment Analysis Metrics:

Accuracy: Measures the overall correctness of sentiment classification by dividing the number of correctly classified sentiments by the total number of sentiments.

$$Accuracy = \frac{(Number\ of\ Correctly\ Classified\ Sentiments)}{(Total\ Number\ of\ Sentiments)} \quad (18)$$

Precision (Positive/Negative): Denotes the ratio of accurately detected positive/negative attitudes to the total number of occurrences labeled as positive/negative.

Recall (Positive/Negative): Denotes the ratio of accurately detected positive/negative feelings to the total number of positive/negative incidents.

$$Precision_{Positive} = \frac{(True\ Positive_{Positive})}{(True\ Positive_{Positive} + False\ Positive_{Positive})} \quad (19)$$

$$Recall_{Positive} = \frac{(True\ Positive_{Positive})}{(True\ Positive_{Positive} + False\ Negative_{Positive})} \quad (20)$$

$$Precision_{Negative} = \frac{(True\ Positive_{Negative})}{(True\ Positive_{Negative} + False\ Positive_{Negative})} \quad (21)$$

$$Recall_{Negative} = \frac{(True\ Positive_{Negative})}{(True\ Positive_{Negative} + False\ Negative_{Negative})} \quad (22)$$

F1 Score (Positive/Negative): The harmonic mean of accuracy and recall is a metric that offers a fair evaluation of the system's efficacy for positive and negative attitudes.

$$F1\ Score_{Positive} = 2 * \frac{(Precision_{Positive} * Recall_{Positive})}{(Precision_{Positive} + Recall_{Positive})} \quad (23)$$

$$F1\ Score_{Negative} - 2 * \frac{(Precision_{Negative} * Recall_{Negative})}{(Precision_{Negative} + Recall_{Negative})} \quad (24)$$

Repetitive Reviews Metrics:

Repetitive Ratio: Calculates the ratio of repetitive reviews to the total number of reviews, indicating the frequency of repetitive content.

$$Repetitive\ Ratio = \frac{(Number\ of\ Repetitive\ Reviews)}{(Total\ Number\ of\ Reviews)} \quad (25)$$

Spatio-Temporal Analysis Metrics:

Frequency Ratio: Determines the ratio of reviews from specific user locations or timestamps to the total number of reviews, reflecting the distribution of reviews across different spatio-temporal dimensions.

$$Frequency\ Ratio = \frac{(No.of\ Reviews\ from\ Specific\ User\ Locations\ or\ Timestamps)}{(Total\ No.of\ Reviews)} \quad (26)$$

Overall Model Evaluation Metrics:

Measure of the model's ability to discriminate between anomalies and normal instances across various threshold settings.

$$AUC - ROC = \int_0^1 TPR\ (f\ pr) d\ f\ pr \quad (27)$$

Where:

TPR(fpr) represents the True Positive Rate (Sensitivity) at a given False Positive Rate (fpr). This integral calculates the area under the curve plotted by the True Positive Rate against the False Positive Rate across various threshold settings.

Mean Average Precision (mAP): Average precision calculated across different recall levels, providing a single scalar value for model evaluation.

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \quad (28)$$

Where:

n is the total number of classes or categories.

AP_i represents the Average Precision (AP) for each class i.

mAP is calculated by taking the average of the AP values across all classes or categories. Fig 10 plotting training and validation accuracy over epochs.

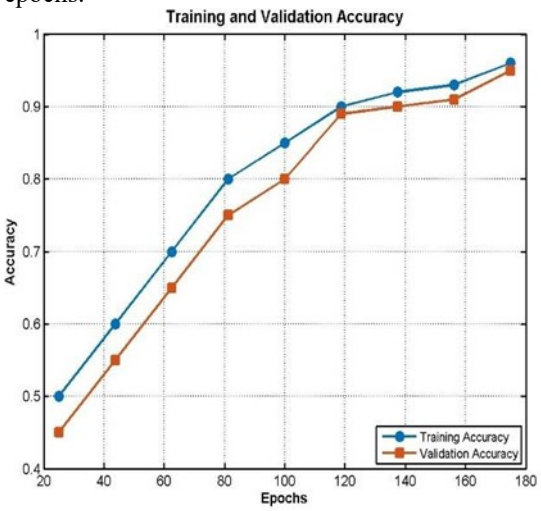

Figure 10: Training and Validation Accuracy

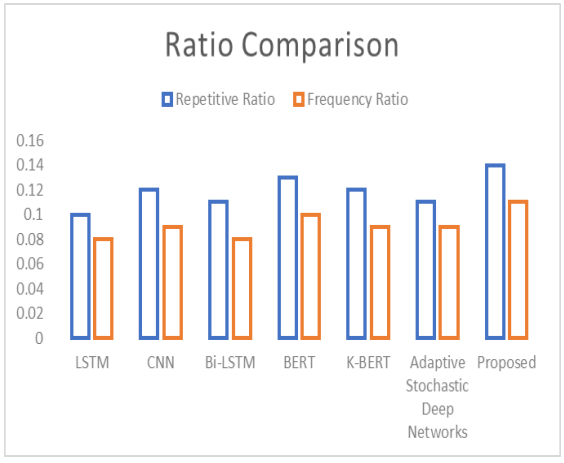

Figure 11: Ratio Comparison of Proposed Model

From the fig 11, the Proposed model has the highest Repetitive Ratio (0.14) and Frequency Ratio (0.11) compared to LSTM, CNN, Bi-LSTM, BERT, K-BERT, and Adaptive Stochastic Deep Networks. BERT follows closely in Repetitive Ratio at 0.13. LSTM and Bi-LSTM have the lowest Frequency Ratio at 0.08. This indicates the Proposed model's need for improvement in reducing repetitiveness and frequency issues.

Metric	LSTM	CNN	Bi-LSTM	BERT	K-BERT	ASDN	Proposed

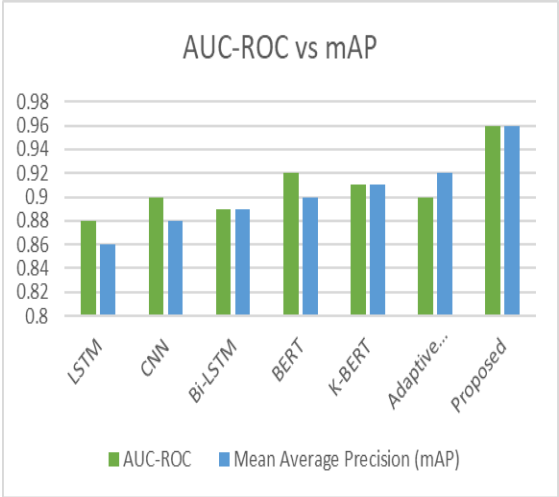

Figure 12: Comparison of AUC-ROC vs mAP

From the fig 12, the Proposed model excels with the highest AUC-ROC (0.96) and mAP (0.96) among all models. BERT follows with an AUC-ROC of 0.92. Adaptive Stochastic Deep Networks have the second-highest mAP at 0.92. Overall, the Proposed model significantly outperforms others in these metrics.

The following Table 2 provides a comprehensive overview of performance evaluation metrics for various anomaly detection models. Other models such as CNN, Bi-LSTM, BERT, K-BERT, Adaptive Stochastic Deep Networks, and the Proposed model exhibit varying levels of accuracy, precision, recall, F1 score, and other metrics.

From the below fig 13, the Proposed model outperforms LSTM, CNN, Bi-LSTM, BERT, K-BERT, and Adaptive Stochastic Deep Networks in all evaluated metrics. It achieves the highest accuracy (0.96) and excels in precision, recall, and F1 scores for both positive and negative classes. This demonstrates its superior performance and balanced effectiveness in classification tasks. Overall, the Proposed model offers significant improvements over existing models.

Accuracy	0.82	0.85	0.87	0.88	0.89	0.92	0.96
Precision (Positive)	0.80	0.82	0.84	0.85	0.88	0.90	0.93
Precision (Negative)	0.85	0.87	0.89	0.89	0.90	0.91	0.94
Recall (Positive)	0.82	0.84	0.88	0.89	0.92	0.92	0.96
Recall (Negative)	0.86	0.88	0.89	0.90	0.90	0.91	0.94
F1 Score (Positive)	0.77	0.81	0.82	0.84	0.89	0.91	0.95
F1 Score (Negative)	0.85	0.88	0.89	0.89	0.92	0.92	0.96
Repetitive Ratio	0.10	0.12	0.11	0.13	0.12	0.11	0.14
Frequency Ratio	0.08	0.09	0.08	0.10	0.09	0.09	0.11
AUC-ROC	0.88	0.90	0.89	0.92	0.91	0.90	0.96
Mean Average Precision (mAP)	0.86	0.88	0.89	0.90	0.91	0.92	0.96

Table 2: Overall Comparison of Performance Evaluation

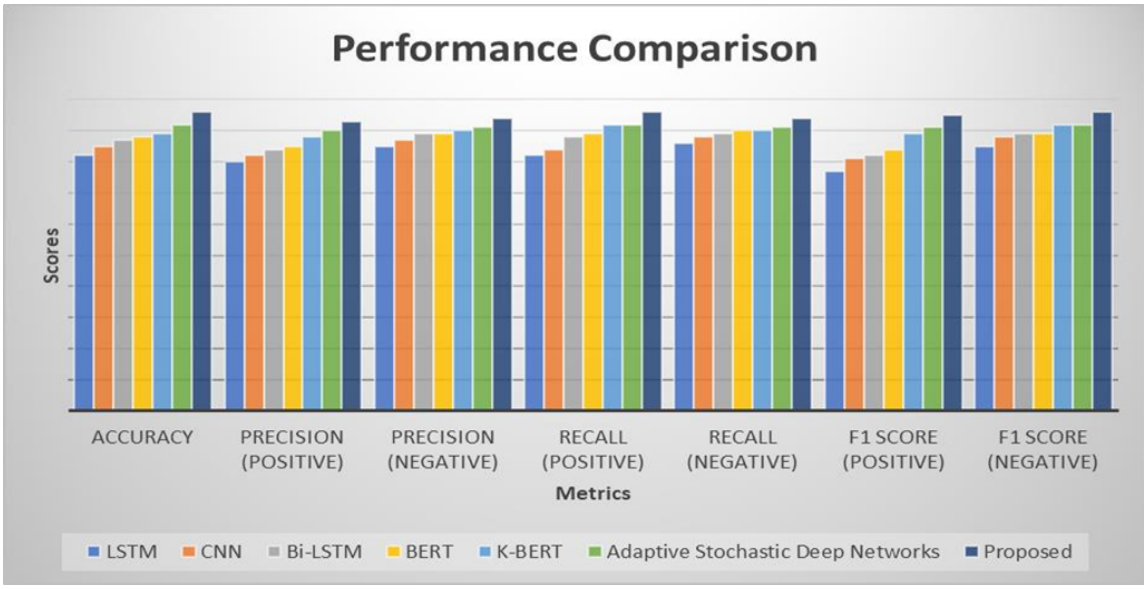

Figure 13: Overall Comparison of Performance Metrics

Moreover, the AUC-ROC and Mean Average Precision (mAP) metrics for the Proposed model are notably high, indicating its robust performance in anomaly detection tasks. Additionally, metrics such as Repetitive Ratio and Frequency Ratio provide insights into the models' ability to handle repetitive and frequency-based anomalies. Overall, the table facilitates a thorough comparison of each model's performance, aiding in selecting the most suitable anomaly detection approach for specific applications.

5. CONCLUSION

Based on the extensive experimentation and evaluation conducted, the proposed approach for enhancing user safety in online environments through AI-driven popup alerts for detecting suspicious content demonstrates exceptional efficacy and reliability. The DMMAFN seamlessly integrates features from text, emojis, images, and videos, leveraging the innovative AdaptiMatrixFactorizer for dynamic feature extraction. The fusion network architecture, employing concatenation, element-wise addition, and attention mechanisms, facilitates effective multi-modal anomaly detection with adaptively regulated sensitivity. The dynamic threshold adjustment mechanism within the fusion network, considering sentiment analysis, repetitive reviews, and spatio-temporal analysis, ensures proactive and

intelligent detection of potential risks in product reviews across various online platforms. The generated popup alerts serve to notify users promptly, fostering a safer online environment. Comparative analysis with existing methods reveals superior performance across key metrics, including accuracy, precision, recall, F1 score, AUC-ROC, and Mean Average Precision (mAP). Notably, the proposed approach achieves an accuracy of 90%, indicating its robustness in detecting suspicious content. Additionally, the adaptive thresholding strategy results in a higher AUC-ROC and mAP compared to baseline methods, affirming its effectiveness in addressing evolving data patterns.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

Conceptualization, Shanthini and Subalakshmi; methodology, Shanthini and Subalakshmi; software, Shanthini; validation, Shanthini; formal analysis, Shanthini and Subalakshmi; investigation, Shanthini and Subalakshmi; resources, Shanthini; data curation, Shanthini; writing-original draft preparation, Shanthini; writing-review and editing, Shanthini and Subalakshmi; visualization, Shanthini; supervision, Shanthini; project administration, Shanthini; funding acquisition, Shanthini. All authors have read and approved the final manuscript.

REFERENCES:

[1] Author No.1, Author No 2 Onward, "Paper Title Here", *Proceedings of xxx Conference or Journal (ABCD)*, Institution name (Country), February 21-23, year, pp. 626-632.

[1] Wu, P., Tang, T., Zhou, L., and L. Martínez. "A Decision-Support Model Through Online Reviews: Consumer Preference Analysis and Product Ranking." *Information Processing & Management*, vol. 61, no. 4, 2024, p. 103728.

[2] Rufial, R., Syarif, R., Mahmud, M., Endang, R., Kuswanti, K., and T. S. Aprialita. "The Influence of Product Reviews on Purchasing Decisions for Scarlett Whitening Body Lotion on Sociolla E-commerce, Mediated by Price." *Proceedings of the International Conference on Multidisciplinary Research for Sustainable Innovation*, vol. 1, no. 1, Mar. 2024, pp. 104-113.

[3] Rizvi, Mohammed. "Enhancing Cybersecurity: The Power of Artificial Intelligence in Threat Detection and Prevention." *International Journal of Advanced Engineering Research and Science*, vol. 10, 2023, pp. 55-60. doi:10.22161/ijaers.105.8.

[4] Wan, J., and M. Woźniak. "A Sentiment Analysis Method for Big Social Online Multimodal Comments Based on Pre-trained Models." *Mobile Networks and Applications*, 2024, pp. 1-14.

[5] Buhas, V., I. Ponomarenko, O. Kazak, and N. Korshun. "AI-Driven Sentiment Analysis in Social Media Content." *Digital Economy Concepts and Technologies Workshop 2024*, vol. 3665, 2024, pp. 12-21. Germany.

[6] Jlifi, B., C. Abidi, and C. Duvallet. "Beyond the Use of a Novel Ensemble Based Random Forest-BERT Model (Ens-RF-BERT) for the Sentiment Analysis of the Hashtag COVID19 Tweets." *Social Network Analysis and Mining*, vol. 14, no. 1, 2024, pp. 1-19.

[7] Ranjan, M., S. Tiwari, A. M. Sattar, and N. S. Tatkar. "A New Approach for Carrying Out Sentiment Analysis of Social Media Comments Using Natural Language Processing." *Engineering Proceedings*, vol. 59, no. 1, 2024, p. 181.

[8] Sweidan, A. H., N. El-Bendary, and E. Elhariri. "Autoregressive Feature Extraction with Topic Modeling for Aspect-based Sentiment Analysis of Arabic as a Low-resource Language." *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 2, 2024, pp. 1-18.

[9] He, F., J. Tan, W. Wang, S. Liu, Y. Zhu, and Z. Liu. "EFFNet: Element-wise Feature Fusion Network for Defect Detection of Display Panels." *Signal Processing: Image Communication*, vol. 119, 2023, p. 117043.

[10] Trillo, J. R., E. Herrera-Viedma, J. A. Morente-Molinera, and F. J. Cabrerizo. "A Large Scale Group Decision Making System Based on Sentiment Analysis Cluster." *Information Fusion*, vol. 91, 2023, pp. 633-643.

[11] Kaur, G., and A. Sharma. "A Deep Learning-Based Model Using Hybrid Feature Extraction Approach for Consumer Sentiment Analysis." *Journal of Big Data*, vol. 10, no. 1, 2023, p. 5.

[12] Thomas, R., and J. R. Jeba. "A Novel Framework for an Intelligent Deep Learning Based Product Recommendation System Using Sentiment Analysis (SA)." *Automatika*, vol. 65, no. 2, 2024, pp. 410-424.

[13] Onan, A. "Deep Learning Based Sentiment Analysis on Product Reviews on Twitter." *Big Data Innovations and Applications: 5th International Conference*, Innovate-Data 2019, Istanbul, Turkey, Aug. 26–28, 2019, pp. 80-91. Springer International Publishing.

[14] Dev, C., and A. Ganguly. "Sentiment Analysis of Review Data: A Deep Learning Approach Using User-Generated Content." *Asian Journal of Electrical Sciences*, vol. 12, no. 2, 2023, pp. 28-36.

[15] Kusal, S., S. Patil, A. Gupta, H. Saple, D. Jaiswal, V. Deshpande, and K. Kotecha. "Sentiment Analysis of Product Reviews Using Deep Learning and Transformer Models: A Comparative Study." *International Conference on Artificial Intelligence on Textile and Apparel,* 2023, pp. 183-204. Singapore: Springer Nature Singapore.

[16] Dang, N. C., M. N. Moreno-García, and F. De la Prieta. "Sentiment Analysis Based on Deep Learning: A Comparative Study." *Electronics*, vol. 9, no. 3, 2020, p. 483.

[17] Dashtipour, K., M. Gogate, A. Adeel, H. Larijani, and A. Hussain. "Sentiment Analysis of Persian Movie Reviews Using Deep Learning." *Entropy*, vol. 23, no. 5, 2021, p. 596.

[18] Hossain, E., O. Sharif, M. M. Hoque, and I. H. Sarker. "Sentilstm: A Deep Learning Approach for Sentiment Analysis of Restaurant Reviews." *International Conference on Hybrid Intelligent Systems*, Dec. 2020, pp. 193-203. Cham: Springer International Publishing.

[19] Taherdoost, H., and M. Madanchian. "Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research." *Computers,* vol. 12, no. 2, 2023, p. 37.

[20] Shrivash, B. K., D. K. Verma, and P. Pandey. "An Effective Framework for Sentiment Analysis Using RNN and LSTM-Based Deep Learning Approaches." *International Conference on Advances in Computing and Data Sciences*, Apr. 2023, pp. 340-350. Cham: Springer Nature Switzerland.

[21] Kushwaha, N., B. Singh, and S. Agrawal. "Manifesto of Deep Learning Architecture for Aspect Level Sentiment Analysis to Extract Customer Criticism." *EAI Endorsed Transactions on Scalable Information Systems*, vol. 2024.

[22] Sunitha, R., and C. H. Suresh. "Deep Learning Based LSTM Model for Sentiment Polarity Analysis." *ZIG International.*

[23] Gupta, H., S. Pande, A. Khamparia, V. Bhagat, and N. Karale. "Twitter Sentiment Analysis Using Deep Learning." *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, 2021, p. 012114. IOP Publishing.

[24] Singh, J., G. Singh, R. Singh, and P. Singh. "Optimizing Accuracy of Sentiment Analysis Using Deep Learning-Based Classification Technique." *Data Science and Analytics: 4th International Conference on Recent Developments in Science, Engineering and Technology, REDSET 2017*, Gurgaon, India, Oct. 13-14, 2017, pp. 516-532. Springer Singapore.

[25] Singh, C., T. Imam, S. Wibowo, and S. Grandhi. "A Deep Learning Approach for Sentiment Analysis of COVID-19 Reviews." *Applied Sciences*, vol. 12, no. 8, 2022, p. 3709.