Journal of Theoretical and Applied Information Technology
15th April 2026. Vol.104. No.7

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

HARNESSING UNCERTAINTY-AWARE AND SCALABLE GMM-RF FOR CROSS-DOMAIN SENTIMENT INTELLIGENCE IN HETEROGENEOUS DIGITAL PLATFORMS

RAMAJAYAM G[1], Dr. VIDYABANU R[2]
[1]Research Scholar, Department of Computer Science, LRG Government Arts College for women, Tirupur, India
[2]Assistant Professor, Department of Computer Science, LRG Government Arts college for women, Tirupur, India
E-mail: [1]ramajayamgurusamy@gmail.com

ABSTRACT

Accurate sentiment classification across diverse online shopping domains remains a complex challenge due to shifting vocabulary, variable sentiment intensity, and non-linear emotional expression. This paper presents a domain-adaptive framework Gaussian Mixture Model-enhanced Random Forest (GMM-RF)—designed to overcome ambiguity and variability in customer reviews by integrating probabilistic clustering with ensemble learning. The GMM component models review as soft sentiment distributions, capturing nuanced emotional overlaps that are often misrepresented by rigid classifiers. These probabilistic encodings guide the Random Forest classifier, which applies cluster-aware bootstrapping, entropy-filtered sampling, and GMM-driven thresholding to form precise and interpretable decision boundaries. Evaluation on Amazon product reviews—Books, DVDs, Electronics, and Kitchen Appliances—demonstrates an average classification accuracy of 95.705%, marking a 7.28% improvement over MMASA. The model also surpasses EBC with a 9.28% gain, confirming its resilience in sentiment-rich, high-entropy environments. GMM-RF requires no domain-specific retraining and is fully scalable, making it ideal for deployment in real-time e-commerce systems. Its design supports adaptive, interpretable sentiment analytics, enabling more reliable customer insight extraction from dynamic, cross-category review streams.

Keywords: *Sentiment Classification, Gaussian Mixture Model, Random Forest, Cross-Domain Analysis, Online Reviews)*

1. INTRODUCTION

Online shopping platforms have reshaped modern commerce by offering customers seamless access to diverse product categories with dynamic user interaction. Consumer reviews posted across these platforms provide not just post-purchase reflections but also serve as primary data streams for market insight extraction [1]. These reviews often contain spontaneous, unstructured, and sentiment-rich narratives influenced by experience, expectation, and perception. The challenge lies in capturing the emotional tone, aspect relevance, and evaluative weight hidden within these expressions. Sentiment analysis enables platforms to extract these subjective cues to drive recommendations, pricing strategies, and quality control processes [2]. As consumer reviews vary in language structure, length, and lexical diversity, automated models must interpret context-sensitive sentiment across numerous review styles, dialects, and sentiment densities without manual intervention [3].

Review content varies significantly between domains like books, electronics, and kitchen appliances, each embedding its own lexical style, evaluation criteria, and sentiment triggers [4]. A sentiment classifier trained on reviews from one domain often performs poorly on another due to changes in vocabulary, expression patterns, and emotional framing. This leads to the emergence of cross-domain opinion mining, which develops generalizable classification models that transfer learned sentiment structures across heterogeneous domains. Such models identify domain-independent indicators and align domain-specific sentiment expressions through techniques like feature projection, pivot-based transfer, and probabilistic dependency modelling [5]. Cross-domain frameworks ensure that sentiment analysis models are scalable, cost-efficient, and effective in diverse

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

e-commerce environments. By minimizing dependence on domain-specific retraining, they enhance deployment feasibility across evolving product categories with minimal annotation overhead [6].

Probabilistic modelling in sentiment classification provides a robust framework for handling ambiguity, sparsity, and context shifts in review data. Models such as Bayesian Networks, Naïve Bayes, and Hidden Markov Models estimate sentiment polarity by computing conditional and joint probabilities over feature sets, capturing linguistic uncertainty and emotional transitions. Bayesian Networks model inter-feature dependencies through directed acyclic graphs, enabling fine-grained representation of sentiment propagation across related terms [7]. Hidden Markov Models capture sequential sentiment shifts within multi-sentence reviews, decoding latent emotional trajectories. Probabilistic classifiers assign calibrated confidence scores to sentiment predictions, supporting risk-aware decision-making and trust-based recommendation systems [8]. These models offer structural interpretability, statistical transparency, and adaptive generalization, making them particularly suitable for high-dimensional, cross-domain sentiment analysis in online shopping. Integrating probabilistic reasoning within sentiment classification pipelines aligns computational predictions with human-like interpretive depth, optimizing the extraction of actionable insights from vast volumes of consumer feedback [9], [10], [11]

1.1. Problem Statement

Online product reviews represent a volatile, sentiment-rich data source critical to digital commerce analysis. Reviews vary across domains in language, expression patterns, and sentiment composition, resulting in significant domain-specific disparities. Traditional classifiers trained on domain-constrained data suffer from low transferability, leading to diminished sentiment detection accuracy when applied across varied product categories such as books, electronics, and appliances. These classifiers often fail to capture ambiguous sentiment transitions and overlapping opinion distributions within user narratives. Moreover, static decision boundaries and rigid label assignments are ineffective in modelling non-linear sentiment spaces where mixed emotion reviews are prevalent. This presents a computational challenge in accurately classifying sentiment under variable linguistic, structural, and contextual review conditions. Models must resolve latent sentiment ambiguity, support domain generalization, and ensure probabilistic interpretability without relying on exhaustive domain-specific retraining or handcrafted lexicons. The above challenges are consistent with limitations identified in existing literature, where current sentiment analysis models struggle to handle overlapping sentiment distributions, domain-specific variability, and non-linear sentiment expressions effectively. These gaps highlight the need for a probabilistic and domain-adaptive framework that can model sentiment uncertainty while maintaining classification robustness across heterogeneous review datasets.

1.2. Motivation

The increasing reliance on user-generated reviews for product evaluation, ranking, and market feedback has created a critical need for robust sentiment classification frameworks. Online shopping ecosystems depend on real-time, context-aware sentiment insights to support recommendation engines, pricing strategies, and vendor reputation metrics. Domain shifts in review vocabulary and emotional framing demand classifiers that adapt across categories without sacrificing interpretability or Precision. Probabilistic sentiment modelling enables the representation of mixed emotional states, facilitating more nuanced classification decisions. Gaussian Mixture Models offer a continuous perspective on sentiment polarity, mitigating the rigidity of conventional hard-label classification. When combined with ensemble learning, these probabilistic distributions can enhance the stability, Precision, and granularity of sentiment classification outcomes. The motivation lies in bridging domain and sentiment ambiguity using hybrid probabilistic-ensemble models that retain interpretability, adaptivity, and computational feasibility in diverse review environments.

1.3. Objective

The primary objective of this research is to develop and validate a Gaussian Mixture Model-enhanced Random Forest (GMM-RF) framework for sentiment classification across online shopping reviews. This framework aims to model continuous sentiment distributions through GMMs and convert probabilistic outputs into robust classification decisions using ensemble-based Random Forests. The methodology is designed to:

- Capture latent sentiment variation through probabilistic cluster membership.
- Encode soft sentiment labels to handle review ambiguity and overlapping emotional zones.

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

- Integrate cluster-aware bootstrapping to improve training data diversity in Random Forest trees.
- Apply probabilistic feature weighting and dependency scoring to refine feature selection.
- Enhance decision tree splitting using GMM-driven thresholds tailored to sentiment density.

The goal is to achieve high precision-recall trade-offs and domain-adaptable performance, enabling scalable deployment of sentiment classification pipelines across structurally heterogeneous and sentiment-diverse review datasets in e-commerce platforms. These objectives are formulated in direct response to the limitations observed in prior studies, aiming to bridge the gap between probabilistic sentiment representation and scalable cross-domain classification.

2. LITERATURE REVIEW

"RMuBERT Sentiment Model" [12] integrates a transformer architecture with self-supervised learning to extract contextual features from Arabic news. It employs Adaptive Modality Contribution and Feature Fusion Learning to dynamically balance linguistic and semantic information. A multi-class sentiment framework processes fused features for classification. The architecture supports dialectal variation through a multi-stage structure that standardizes representation across Arabic modalities using text encoding and sentiment alignment mechanisms. "Sentiment Analysis Overview" [13] categorizes sentiment analysis methods into lexicon-based, machine learning, and hybrid approaches. It presents structured mechanisms for feature extraction, polarity detection, and sentiment classification across domains. It also outlines methods for lexicon generation, rule-based tagging, and statistical sentiment modelling in structured pipelines. The taxonomy maps models to application-specific analysis tasks using distinct linguistic or statistical inference strategies. "Modality Completion Sentiment" [14] generates full-modality samples to address missing data in multi-modal sentiment analysis. A similarity-based mechanism reconstructs absent modalities. A transformer encoder aligns the reconstructed features with original ones to ensure representation consistency. Decision-level fusion aggregates sentiment outputs from all modalities, allowing classification to proceed under incomplete input conditions using aligned feature vectors. "Sustainability Sentiment Study" [15] extract sentiment signals from sustainability-related social media text. Feature extraction techniques isolate environmental terms and sentiment-bearing tokens. Structured classification modules assign polarity labels based on sentiment lexicon scores or neural representations. The system processes real-time streams and text batches to identify sentiment linked to sustainability keywords, topical mentions, or policy terms.

"Sacred Text Sentiment" [16] processes sacred texts to extract emotional tones. Preprocessing includes tokenization and Part-of-Speech tagging for adjective and verb identification. A sentiment lexicon assigns polarity scores to individual terms. Temporal and thematic sentiment trajectories are mapped using aggregated word-level polarity signals. Cross-text comparison methods align linguistic structures across religious scripts for sentiment pattern analysis. "Disentangled Fusion Sentiment" [17] disentangles unimodal and joint features. A distribution constraint module separates modality-specific and fused representations. A hybrid loss function handles inductive bias, signal reconstruction, and semantic consistency. The fusion system uses probabilistic distribution control to maintain feature independence. Each modality contributes separately before final sentiment decision-making via disentangled fusion layers. "AI News Sentiment Study" [18] extracts dominant themes from news articles topic modelling using Latent Dirichlet Allocation. Sentiment analysis is performed using rule-based polarity assignment on headlines. Cross-national datasets are segmented using metadata, and textual features are mapped into sentiment polarity via keyword frequency and syntactic analysis. Timeline-based sentiment shifts are tracked using structured sentiment aggregation and temporal segmentation. "Vision-Language Sentiment" [19] links visual and textual data in multi-modal aspect-based sentiment analysis. Scene graphs extract visual elements, and syntactic parsers process text inputs. Structural alignment maps image regions to dependency structures. Semantic alignment links word phrases to visual segments. A cross-modal attention matrix refines sentiment-relevant features, enabling aspect-level polarity classification from fused multi-modal embeddings.

"Fine-Grained Multi-modal Sentiment" [20] applies a graph convolutional network to model syntactic dependencies in text. Visual features are extracted using adjective-noun pair analysis from images. A fusion mechanism combines weighted textual and visual elements dynamically. A selfattention layer refines feature contributions from both modalities. A joint loss function enables concurrent learning of aspect detection and sentiment polarity assignment. "Socio-Pragmatic Code-Switching" [21] combines XLM-RoBERTa with a DistilBERT-based socio-pragmatic embedding layer to process Dravidian–English code-switched data. Language alternation points are detected using syntactic tagging. Dual-channel attention separates sentiment indicators derived from lexical and sociolinguistic patterns. Adversarial training and multi-level feature extraction process sentiment transitions and identify polarity using code-switch-sensitive representation learning. "Multi-Level Sentiment Fusion" [22] combines textual, visual, and auditory data using hierarchical feature extraction. A Scale-Global Information Extraction module captures sentiment variation across levels. A memory module retains spatial sentiment features. A two-step cross-attention mechanism fuses inter-modal signals. Contextual weighting adjusts the fusion based on sentiment intensity, forming a unified sentiment representation. "Diffusion-Based Sentiment Model" [23] reconstruct missing sentiment cues across text, image, and audio inputs. A machine reading comprehension module identifies contextually linked sentiment components. Diffusion models iteratively refine feature representations. A hybrid alignment module maps multi-modal embeddings into a shared semantic space. A transformer encoder extracts sentiment signals from denoised representations.

"Entropy-Based Classifier (EBC)" [24] employs a semi-supervised learning framework for cross-domain sentiment analysis using modified entropy weighting and bipartite graph clustering. It begins with domain-independent lexicon construction and then propagates sentiment polarity to domain-specific terms through contextual associations. The method offers interpretability and lightweight execution, making it effective for structured opinion mining. However, it heavily relies on frequency-based heuristics, struggles with sparse or ambiguous data, and cannot model overlapping sentiment expressions, limiting its scalability across sentiment-dense or linguistically diverse review environments. "Multi-Modal Aspect-Based Sentiment Analysis (MMASA)" [25] leverages both text and image data using multi-modal interaction layers and adversarial alignment to detect aspect-specific sentiments. It applies Bi-LSTM and ResNet encoders, aligning embeddings through a discriminator network to unify sentiment space across modalities. The model excels in rich visual-text contexts, offering fine-grained aspect detection. Despite its accuracy, MMASA is computationally intensive, sensitive to modality imbalance, and lacks adaptability in purely textual or domain-shifting environments. These constraints reduce its effectiveness in large-scale, text-heavy online shopping platforms.

Recent studies have explored the application of machine learning techniques for pattern detection and classification in complex data environments. A machine learning-based approach has been used to analyze cybercrime patterns, demonstrating the effectiveness of data-driven models in identifying hidden structures within unstructured datasets [26]. In the context of sentiment analysis, a rejuvenated Particle Swarm Optimization (PSO)-based classifier has been proposed for large-scale sentiment data, improving feature selection and classification efficiency in high-dimensional spaces [27]. Similarly, a renewed Genetic Algorithm-based framework has been applied for cross-domain sentiment classification, addressing domain variability and enhancing generalization across heterogeneous datasets [28].

Beyond classification, secure and efficient data handling has also been explored through fuzzy-identity-based encryption mechanisms for data sharing [29], highlighting the importance of reliable data management in intelligent systems. In addition, AI-driven models have been increasingly utilized for decision-making and business analytics, emphasizing the role of scalable and adaptive learning frameworks in real-world applications [30]. Broader perspectives on AI integration, such as cultural intelligence in education systems, further underline the expanding scope of intelligent systems across domains [31].

In networked environments, adaptive routing protocols have been developed to enhance connectivity and quality of service, reflecting the need for efficient and dynamic system design in data-driven applications [32]. Foundational works in bio-inspired intelligence and optimization have further contributed to smart decision-making frameworks by enabling adaptive search and optimization mechanisms in complex problem spaces [33]–[36].

While these studies demonstrate advancements in optimization, classification, and intelligent decision-making, they do not explicitly address sentiment uncertainty and overlapping sentiment distributions in cross-domain textual data. In contrast, the proposed GMM-RF framework integrates probabilistic sentiment modelling with ensemble learning to effectively capture ambiguity and improve sentiment classification performance across heterogeneous domains.

Despite the diversity of existing sentiment analysis approaches, several limitations remain evident. Transformer-based and multimodal models demonstrate strong contextual learning but often require high computational resources and domain-specific tuning, limiting scalability across heterogeneous review datasets. Entropy-based and lexicon-driven methods offer interpretability but fail to capture overlapping sentiment expressions and non-linear sentiment transitions. Furthermore, many models assume rigid sentiment boundaries, which reduces their effectiveness in handling mixed-emotion reviews common in e-commerce platforms. These limitations highlight the need for a probabilistic and scalable framework that can model sentiment uncertainty while maintaining classification robustness across domains. The proposed GMM-RF framework is designed to address these gaps by integrating probabilistic clustering with ensemble learning to support cross-domain sentiment generalization.

3. GAUSSIAN MIXTURE MODEL-ENHANCED RANDOM FOREST

This section introduces the proposed hybrid sentiment classification framework designed to address the limitations of existing models in handling domain diversity, sentiment overlap, and feature ambiguity in online reviews. By integrating Gaussian Mixture Models with Random Forest classifiers, the framework captures soft sentiment boundaries through probabilistic clustering and enforces robust classification using ensemble-based decision strategies. This section outlines the structural innovation in transforming sentiment-rich text into interpretable probability spaces, followed by adaptive feature encoding and entropy-regulated instance selection. The framework aims to enhance accuracy, generalization, and interpretability, offering a scalable solution for real-world sentiment analysis across product domains.

3.1. Capturing Latent Sentiment Distributions in Online Shopping Reviews using GMM

The Gaussian Mixture Model (GMM) has provided a robust probabilistic framework for capturing latent sentiment structures in online shopping reviews. Sentiment expressions in e-commerce platforms exhibit complex variations due to differences in customer preferences, review styles, product categories, and sentiment intensities. Conventional sentiment analysis models have struggled handling such diverse sentiment representations due to fixed decision boundaries, leading to performance bottlenecks in precision-recall-oriented sentiment classification. The integration of GMM within Random Forest (RF) has enabled an optimized sentiment representation by modelling the continuous distribution of sentiments, thereby refining the feature transformation process.

The sentiment space has exhibited non-linear variations where a single classification boundary has remained insufficient to capture complex emotional expressions in online reviews. GMM has mapped the sentiment distribution to a mixture of Gaussian components, allowing a continuous, probabilistic representation of sentiments. The optimized modelling approach has ensured that overlapping sentiments, such as “slightly positive” and “highly positive,” have been distinguished with high granularity. The probability density function (PDF) of the GMM for sentiment representation has been given by:

$$P(s) = \sum_{k=1}^{K} \pi_k \cdot N(s|\mu_k, \Sigma_k) \qquad (1)$$

where $P(s)$ has denoted the probability density of a sentiment feature s, K has represented the total number of Gaussian components, π_k has been a mixture of satisfying $\sum_{k=1}^{K} \pi_k = 1$, and $N(s|\mu_k, \Sigma_k)$ has represented the Gaussian distribution with mean μ_k and covariance Σ_k.

The optimized GMM-based sentiment mapping has provided a fine-grained probability representation of review sentiment scores. Unlike hard-classified sentiment labels, GMM has allowed adaptive sentiment labelling, making classification highly responsive to minor variations in textual features. The continuous mixture density estimation has refined feature transformation in the RF model, leading to increased classification robustness.

Expectation-Maximization (EM) has been employed to optimize sentiment probability estimation by refining the parameters of the Gaussian distributions iteratively. The EM algorithm has maximized the likelihood of the observed sentiment data by updating parameters in Expectation (E-step) and Maximization (M-step).

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

The likelihood function for sentiment feature distribution estimation has been given as follows:

$$L(\Theta|S) = \prod_{i=1}^{N}\sum_{k=1}^{K}\pi_k \cdot N(s_i|\mu_k, \Sigma_k) \quad (2)$$

where $L(\Theta|S)$ has represented the likelihood of sentiment data given the parameters $\Theta = \{\pi_k, \mu_k, \Sigma_k\}$, and N has been the number of sentiment instances.

The E-step has computed the responsibility γ_{ik}, which has denoted the probability that the i-th review has belonged to the k-th Gaussian component:

$$\gamma_{ik} = \frac{\pi_k \cdot N(s_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K}\pi_k \cdot N(s_i|\mu_k, \Sigma_k)} \quad (3)$$

The M-step has updated the parameters based on the computed responsibilities:

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^{N}\gamma_{ik}s_i}{\sum_{i=1}^{N}\gamma_{ik}} \quad (4)$$

$$\sum\nolimits_k^{(t+1)} = \frac{\sum_{i=1}^{N}\gamma_{ik}(s_i-\mu_k)(s_i-\mu_k)^T}{\sum_{i=1}^{N}\gamma_{ik}} \quad (5)$$

After optimizing sentiment distributions, feature encoding has been enhanced by incorporating soft cluster assignments. Each review has been assigned a probability distribution across multiple Gaussian components rather than a fixed sentiment label. This process has eliminated the rigid classification structure, making RF more adaptable to review complexity. The probability-based feature transformation has been defined as:

$$\tilde{s}_i = \sum_{k=1}^{K}\gamma_{ik} \cdot \mu_k \quad (6)$$

where $\tilde{s}_i$ has denoted the transformed sentiment representation for the i-th review, integrating the weighted sum of Gaussian means based on the posterior probabilities γ_{ik}.

This feature transformation has allowed sentiment blending, where mixed-emotion reviews have been encoded with more interpretability. The transition from categorical sentiment representation to a GMM-refined probabilistic space has significantly improved precision-recall trade-offs in RF classification.

An optimized cluster selection mechanism has been employed to refine the model further. Reviews with high uncertainty in cluster assignment have undergone entropy-based filtering, ensuring that ambiguous instances do not introduce classification noise. The entropy of cluster assignments has been computed as:

$$H(s_i) = -\sum_{k=1}^{K}\gamma_{ik} log\, \gamma_{ik} \quad (7)$$

where $H(s_i)$ has measured the uncertainty in sentiment representation. Reviews exceeding a predefined entropy threshold have been re-weighted before passing into the RF model.

3.2. Enhanced Feature Representation Using Probabilistic Clustering

Extracting latent sentiment structures has improved understanding of contextual variations in online shopping reviews. However, an optimized feature representation has remained necessary to bridge the gap between raw sentiment attributes and their interpretability in classification models. The incorporation of probabilistic clustering has structured sentiment data into a refined feature space where review characteristics have been transformed based on probability-weighted assignments. The refined representation has improved the precision-recall trade-off in classification, ensuring that even subtle sentiment nuances have been captured efficiently.

Feature transformation has played a crucial role in ensuring that the sentiment model has retained meaningful distinctions between overlapping sentiment expressions. Traditional feature extraction techniques have relied on hard-coded representations, leading to rigid boundaries between positive, neutral, and negative sentiments. To overcome this limitation, probabilistic clustering has ensured that each sentiment instance has been represented as a mixture of probability distributions rather than a single categorical value. The transformation function has been represented as follows:

$$\tilde{X}_i = \sum_{k=1}^{K}P(C_k|x_i) \cdot \phi_k \quad (8)$$

where $\tilde{X}_i$ has denoted the transformed feature vector for the i-th sentiment instance, $P(C_k|x_i)$ has represented the posterior probability of the i-th sentiment belonging to the k-th Gaussian component, and ϕ_k has indicated the centroid representation of the k-th sentiment cluster.

This approach has provided an optimized transformation, where sentiment data points have been assigned fractional cluster memberships rather than discrete classifications. The probabilistic distribution of sentiments has improved the granularity of feature encoding, making the Random Forest model more adaptable to contextual variations in online shopping reviews.

Each sentiment component has been assigned a weight based on the degree of association with a corresponding cluster to strengthen feature interpretability. This probability-weighted representation has allowed the model to assign higher importance to dominant sentiment patterns while retaining minor variations for subtle classifications. The probability-weighted feature encoding has been formulated as follows:

$$Z_i = \sum_{k=1}^{K} w_k \cdot P(C_k|x_i) \cdot f_k \tag{9}$$

where Z_i has represented the transformed feature space for the i-th sentiment review, w_k has been the cluster-specific weight assigned to the k-th Gaussian component, $P(C_k|x_i)$ has denoted the probability of sentiment instance x_i belonging to the k-th cluster, and f_k has indicated the optimized feature centroid for the k-th cluster.

A key challenge in probabilistic feature representation has been the high-dimensional nature of transformed sentiment attributes. Sentiment Variance Analysis (SVA) has been applied to reduce redundant dimensions while preserving discriminative power in the feature space to mitigate this issue. The variance-based feature optimization has been represented as:

$$V_j = \frac{1}{N} \sum_{i=1}^{N} P(C_k|x_i) \cdot \left(x_{ij} - \mu_j\right)^2 \tag{10}$$

where V_j has denoted the variance of the j-th feature, N has been the number of sentiment instances, x_{ij} has represented the feature value for the i-th review in the j-th dimension, and μ_j has indicated the mean of the j-th feature across all reviews.

This technique has ensured that highly informative features have been retained, whereas dimensions contributing to marginal variations have been eliminated. The optimized feature space has improved classification robustness, enabling Random Forest to focus on high-impact sentiment differentiators.

Online shopping reviews have exhibited inherent variability, where similar sentiment expressions have appeared across multiple product categories. The presence of category-dependent sentiment shifts has required an optimized refinement mechanism to adapt feature representation accordingly. To achieve this, the Cluster-Specific Sentiment Refinement (CSSR) process has been implemented, ensuring that sentiment categories have been aligned dynamically based on their distributional properties. The refinement function has been structured as:

$$S_i^{opt} = \sum_{k=1}^{K} \alpha_k \cdot P(C_k|x_i) \cdot S_k \tag{11}$$

where S_i^{opt} has denoted the refined sentiment representation for the i-th review, α_k has been an adaptive tuning parameter optimizing cluster influence, and S_k has represented the sentiment centroid of the k-th cluster.

3.3. Optimized Bootstrapping with Cluster-Aware Sampling

Bootstrapping has been a fundamental process in Random Forest (RF) training, ensuring diversity across decision trees by generating resampled datasets. However, conventional bootstrapping has introduced bias when dealing with imbalanced sentiment distributions in online shopping reviews. Integrating Gaussian Mixture Model (GMM) clustering into the bootstrapping process has ensured an optimized sampling strategy, where data has been proportionally selected based on cluster distributions. This approach has retained sentiment diversity, preventing overfitting while improving classification precision and recall.

Traditional bootstrapping has randomly sampled instances, leading to imbalanced training subsets, particularly in datasets where certain sentiment classes have been underrepresented. Applying Cluster-Aware Sampling (CAS) has ensured that each sampled subset has preserved the natural sentiment distributions, allowing decision trees to receive a balanced mix of sentiment instances. The optimized resampling probability function has been defined as:

$$P(x_i) = \frac{P(C_k|x_i)}{\sum_{j=1}^{N} P(C_k|x_j)} \tag{12}$$

where $P(x_i)$ has represented the probability of selecting the i-th sentiment instance, $P(C_k|x_i)$ has denoted the probability that the i-th sentiment belongs to the k-th Gaussian cluster, and N has been the total number of sentiment instances.

An optimized instance selection mechanism has been incorporated to further refine the bootstrapping process, where sentiment instances have been sampled proportionally to their cluster significance. This process has provided greater adaptability by ensuring highly representative instances have been assigned higher probabilities during sampling iterations. The instance selection function has been structured as:

$$S_k = \frac{1}{|C_k|} \sum_{i \in C_k} P(x_i) \cdot w_i \tag{13}$$

where S_k has denoted the cluster-specific sampling probability for the k-th sentiment cluster, $|C_k|$ has represented the total number of instances in the k-th cluster, $P(x_i)$ has been the probability of selecting the i-th instance, and w_i has indicated the probabilistic weight assigned to the instance.

Once the resampling probabilities have been established, each bootstrapped subset will be generated dynamically while ensuring that each tree in the RF model receives balanced sentiment representations. Unlike traditional bootstrapping, where duplicate instances have often been randomly selected, the optimized process has regulated instance repetition, ensuring that each subset has captured distinct sentiment nuances. The optimized subset function has been defined as:

$$B_t = \{x_i | x_i \sim P(x_i) \; and \; |B_t| = M\} \tag{14}$$

where B_t has denoted the bootstrapped subset for the t-th decision tree, $P(x_i)$ has been the resampling probability derived from GMM-based cluster weights, and M has represented the desired subset size.

3.4. Refined Feature Importance Through Probabilistic Weighting

The application of feature importance refinement has significantly impacted the performance of sentiment classification models by ensuring that the most discriminative attributes have been prioritized during Random Forest (RF) training. Conventional feature selection techniques have relied on frequency-based measures, which cannot capture context-dependent sentiment variations. Integrating Gaussian Mixture Model (GMM)-driven probabilistic weighting has ensured that each sentiment feature has been assigned a dynamic importance score, improving classification precision and recall.

A probabilistic weighting mechanism has been essential in ensuring that sentiment features have been differentiated based on their predictive significance. Instead of assigning uniform weights, the refined method has incorporated posterior probabilities derived from GMM clustering, ensuring that each feature's impact has been adjusted dynamically based on its association with sentiment clusters. The probabilistic feature weight function has been defined as:

$$w_j = \sum_{k=1}^{K} P(C_k|x_j) \cdot \delta_k \tag{15}$$

where w_j has represented the importance weight of the j-th feature, $P(C_k|x_j)$ has denoted the probability of the j-th feature belonging to the k-th cluster, and δ_k the cluster-specific weighting factor has ensured that high-relevance clusters have contributed more to the feature importance score.

Once feature weights have been computed, a ranking-based approach has been implemented to select high-impact sentiment features while removing noisy or redundant attributes. Instead of using fixed thresholds, an adaptive probability-based ranking mechanism has been employed to ensure that feature selection has been context-sensitive. The probability-based ranking function has been formulated as:

$$R_j = \frac{w_j}{\sum_{m=1}^{M} w_m} \tag{16}$$

where R_j has represented the normalized ranking score for the j-th feature, w_j has been the computed probabilistic weight, and M has denoted the total number of features.

Sentiment expressions in online shopping reviews have exhibited non-independent relationships, where certain features have influenced each other in determining sentiment polarity. Traditional feature selection methods have failed to capture such dependencies, leading to suboptimal classification performance. A dependency-aware feature optimization mechanism has been integrated to address this, ensuring that correlated features have been grouped adaptively based on their mutual probability distribution. The dependency function has been structured as:

$$D_{ij} = \sum_{k=1}^{K} P\left(C_k | x_i, x_j\right) \cdot \lambda_k \quad (17)$$

where D_{ij} has represented the dependency score between features i and j, $P\left(C_k | x_i, x_j\right)$ has been the joint probability of features i and j belonging to the k-th sentiment cluster, and λ_k has been a cluster-specific tuning parameter.

A common issue in sentiment classification has been overfitting, where excessive low-impact features have led to overly complex decision boundaries. Implementing Cluster-Weighted Feature Pruning (CWFP) has mitigated this issue by systematically removing weakly contributing features while ensuring that the model's generalization capability has been preserved. The pruning function has been given as:

$$F_{pruned} = \{x_j | R_j < \tau\} \quad (18)$$

where F_{pruned} has denoted the set of pruned features, R_j has been the ranking score of the j-th feature, and τ has represented a dynamically adjusted pruning threshold.

Once probabilistic feature selection has been completed, a dynamic scaling mechanism has been applied to ensure that feature values have been normalized based on their probabilistic distribution. Instead of using fixed scaling factors, this approach has ensured that each feature's contribution has been aligned with its sentiment classification potential. The scaling function has been structured as:

$$X_j^{scaled} = \frac{X_j - \mu_j}{\sigma_j} \times w_j \quad (19)$$

where X_j^{scaled} has represented the scaled feature value, X_j has denoted the original feature value, μ_j has been the mean value of the j-th feature, σ_j has represented the standard deviation, and w_j has been the computed probabilistic weight.

3.5. Improved Decision Tree Splitting Using GMM-Driven Thresholds

Decision tree splitting has played a crucial role in determining the classification accuracy of Random Forest (RF) models. Traditional decision tree-based classifiers have relied on impurity reduction criteria, such as the Gini index or entropy, to select optimal split points. However, these methods have not effectively accounted for complex sentiment variations, leading to suboptimal decision boundaries. Integrating Gaussian Mixture Model (GMM)-driven thresholds has ensured that split points have been selected based on probabilistic feature distributions, enhancing precision-recall trade-offs in sentiment classification for online shopping reviews.

Conventional tree-splitting techniques have determined decision boundaries based on global feature distributions, which have not effectively captured local sentiment structures. By incorporating GMM-driven probabilistic thresholding, each split point has been dynamically adjusted based on the sentiment cluster density, ensuring an optimized feature partitioning strategy. The GMM-enhanced split threshold function has been defined as:

$$T_j = \sum_{k=1}^{K} P(C_k | X_j) \cdot \theta_k \quad (20)$$

where T_j has represented the optimized split threshold for the j-th feature, $P(C_k | X_j)$ has been the probability of the j-th feature belonging to the k-th Gaussian sentiment cluster, and θ_k has indicated the cluster-specific decision boundary.

The selection of optimal feature splits has required the minimization of classification impurity, ensuring that each split has maximized sentiment class separation. Instead of relying solely on Gini impurity or entropy, a cluster-weighted impurity reduction function has been applied, refining the decision boundary selection process. The impurity reduction function has been formulated as:

$$I_{split} = \sum_{k=1}^{K} w_k \cdot [H(P_L) + H(P_R)] \quad (21)$$

where I_{split} has represented the cluster-weighted impurity reduction, w_k has been the importance weight of the k-th sentiment cluster, and $H(P_L)$ and $H(P_R)$ have denoted the entropy of the left and right partitions after the split.

Sentiment-based decision boundaries have often exhibited high variability, where certain sentiment categories have been underrepresented in tree splits. An adaptive partitioning strategy has been implemented to mitigate this, ensuring that each split

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

has preserved sentiment diversity while avoiding overfitting to dominant sentiment classes. The adaptive partitioning function has been defined as:

$$S_{opt} = \frac{1}{K}\sum_{k=1}^{K} \alpha_k \cdot P(C_k|X) \quad (22)$$

where S_{opt} has represented the optimized sentiment partition, α_k has been a cluster-specific weight adjustment parameter, and $P(C_k|X)$ has denoted the probability of a sentiment feature belonging to cluster k.

A key limitation in traditional tree-based classifiers has been the inability to prioritize features dynamically based on their sentiment discriminative power. To address this issue, a cluster-sensitive feature selection process has been incorporated, where sentiment feature importance scores have been computed before tree node splitting. The feature selection function has been structured as:

$$F_{split} = \sum_{j=1}^{M} P\left(X_j|C_k\right) \cdot \beta_j \quad (23)$$

where F_{split} has represented the selected feature for node splitting, $P\left(X_j|C_k\right)$ has denoted the probability of the j-th feature belonging to the k-th sentiment cluster, and β_j has been the feature-specific selection weight.

Overfitting has remained a persistent challenge in decision tree-based models, where excessive tree depth has led to classification instability. By integrating GMM-driven threshold-based pruning, tree depth has been regulated dynamically, ensuring that overfitting risks have been minimized while classification generalization has been maintained. The depth control function has been given as:

$$D_{opt} = \sum_{k=1}^{K} \gamma_k \cdot log_2(\theta_k) \quad (24)$$

where D_{opt} has denoted the optimized tree depth, γ_k has been a cluster sensitivity parameter, and θ_k has represented the cluster-specific threshold adjustment factor.

3.6. Enhanced Model Robustness by Filtering Noisy Sentiment Data

The classification of online shopping reviews has been significantly affected by noisy sentiment data, where ambiguous, contradictory, or misclassified reviews have reduced the performance of sentiment classifiers. Traditional noise reduction techniques have often relied on heuristic filtering, failing to effectively capture sentiment inconsistencies. Integrating Gaussian Mixture Model (GMM)-based noise filtering has ensured that outliers and ambiguous sentiment instances have been identified and handled systematically, improving Random Forest (RF) classification robustness.

The presence of outliers in sentiment data has introduced classification bias, where misclassified sentiment instances have distorted feature distributions. A probabilistic noise detection function has been employed to mitigate this, ensuring that low-confidence sentiment samples have been re-weighted or removed dynamically.The outlier detection function has been structured as:

$$O_i = 1 - \sum_{k=1}^{K} P(C_k|x_i) \quad (25)$$

where O_i has represented the outlier score for the i-th sentiment instance, $P(C_k|x_i)$ has denoted the probability of the i-th instance belonging to the k-th sentiment cluster, and K has been the total number of sentiment clusters.

Once outliers have been detected, a cluster-based thresholding mechanism has been applied to determine whether a noisy instance should be re-weighted or discarded. Instead of a fixed rejection threshold, an adaptive variance thresholding mechanism has been implemented to ensure that sentiment variations are handled contextually. The variance-based noise threshold function has been given as:

$$T_k = \sigma_k^2 + \eta \cdot \mu_k \quad (26)$$

where T_k has represented the adaptive threshold for the k-th sentiment cluster, σ_k^2 has denoted the variance of sentiment feature distributions within the k-th cluster, μ_k has been the mean sentiment score, and η has been a tuning parameter controlling noise sensitivity.

Not all noisy instances have been purely erroneous classifications; some reviews have exhibited contextual ambiguity where multiple sentiment classes have overlapped. A probability-based re-weighting function has been applied to handle such cases, ensuring that low-confidence

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

reviews have been down-weighted rather than removed entirely. The probabilistic re-weighting function has been defined as:

$$W_i = \sum_{k=1}^{K} P(C_k|x_i) \cdot \gamma_k \tag{27}$$

where W_i has represented the new weight for the i-th sentiment instance, $P(C_k|x_i)$ has been the posterior probability of sentiment cluster membership and γ_k has been a cluster-specific importance weighting factor.

Beyond instance-based filtering, noisy sentiment data has also affected feature distributions, where certain sentiment attributes have exhibited inconsistent importance scores. A cluster-based feature refinement approach has been implemented to address this, ensuring that features most affected by noisy sentiment data have been adjusted adaptively. The feature refinement function has been structured as:

$$F_j^{adj} = F_j - \sum_{i=1}^{N} O_i \cdot X_{ij} \tag{28}$$

where F_j^{adj} has represented the adjusted feature importance score, F_j has been the original importance score, O_i has denoted the outlier score of the i-th sentiment instance, and X_{ij} has been the value of the j-th feature in the i-th review.

High entropy scores have often reflected sentiment data uncertainty, where reviews have exhibited equal probabilities across multiple sentiment clusters. An entropy-based filtering mechanism has been applied to address this issue, ensuring that high-uncertainty sentiment instances have been either re-weighted or excluded based on their entropy scores. The entropy function has been given as:

$$H(x_i) = \sum_{k=1}^{K} P(C_k|x_i) log P\,(C_k|x_i) \tag{29}$$

where $H(x_i)$ has represented the entropy of the i-th sentiment instance, and $P(C_k|x_i)$ has denoted the probability of sentiment cluster membership.

3.7. Incorporation of Cluster-Weighted Voting for Improved Sentiment Prediction

The classification of online shopping reviews has relied on ensemble learning techniques, where multiple decision trees have contributed to the final sentiment prediction. Traditional Random Forest (RF) voting mechanisms have followed a simple majority rule, which has not adequately accounted for the confidence of individual decision trees. Incorporating Gaussian Mixture Model (GMM)-driven cluster-weighted voting has ensured that predictions have been aggregated adaptively, allowing higher confidence predictions to contribute more significantly to the final sentiment classification.

A major limitation of majority voting has been the equal treatment of all decision trees, leading to cases where low-confidence predictions have influenced the final classification. A probability-driven voting mechanism has been employed to address this, ensuring that predictions have been weighted based on their alignment with sentiment cluster distributions. The cluster-weighted voting function has been given as:

$$V_c = \sum_{t=1}^{T} P(C_k|X_t) \cdot d_t \tag{30}$$

where V_c has represented the final sentiment vote for class c, T has been the total number of decision trees, $P(C_k|X_t)$ has denoted the probability of sentiment cluster k influencing the t-th decision tree, and d_t has been the classification output of the t-th decision tree.

Not all sentiment clusters have contributed equally to classification decisions, as certain clusters have exhibited greater separability than others. A confidence-based sentiment weighting function has been applied to optimize this process, ensuring that high-confidence clusters have been prioritized during voting. The adaptive weighting function has been structured as:

$$W_k = \frac{1}{\sigma_k^2 + \epsilon} \tag{31}$$

where W_k has represented the confidence weight of sentiment cluster k, σ_k^2 has been the variance of sentiment feature distributions in the cluster, and ϵ has been a regularization parameter preventing numerical instability.

Online shopping reviews have often exhibited polarized sentiment distributions, where certain sentiment categories have been more frequent than others. A probabilistic sentiment score normalization function has been applied to ensure balanced classification, adjusting class probabilities based on cluster-specific priors. The normalization function has been defined as:

$$S_c = \frac{V_c}{\sum_{j=1}^{C} V_j} \tag{32}$$

where S_c has represented the normalized sentiment score for class c, V_c has been the raw sentiment vote, and C has denoted the total number of sentiment categories.

Since decision trees have individually contributed to sentiment classification, certain cases have required additional refinement, where predictions have conflicted due to cluster-level disagreements. A refinement function has been applied to handle this issue, ensuring that the final sentiment decision has aligned with GMM-derived sentiment distributions. The decision refinement function has been structured as:

$$D_f = \arg\max_c \sum_{k=1}^{K} P(C_k|X) \cdot S_c \tag{33}$$

where D_f has represented the final sentiment classification, $P(C_k|X)$ has been the GMM-derived probability of sentiment cluster k, and S_c has denoted the normalized sentiment score for class c.

Certain sentiment predictions have exhibited systematic misclassification patterns, particularly when specific clusters have been consistently confused with similar sentiment classes. A misclassification handling mechanism has been introduced to mitigate this issue, ensuring that misclassified predictions have been weighted accordingly. The misclassification weight function has been given as:

$$M_c = \sum_{k=1}^{K} P(C_k|X) \cdot E_k \tag{34}$$

where M_c has represented the misclassification penalty score for class c, E_k has been the historical misclassification rate of cluster k, and $P(C_k|X)$ has denoted the GMM-derived probability of cluster membership.

3.8. Implementation of Adaptive Sentiment Classification Thresholds

Traditional sentiment classification techniques have employed fixed decision thresholds, leading to rigid classification boundaries that have failed to account for contextual variations in sentiment expressions. Integrating Gaussian Mixture Model (GMM)-driven adaptive thresholds has ensured that classification decisions have been dynamically adjusted based on sentiment distribution properties, improving precision-recall balance in online shopping sentiment analysis.

Fixed classification thresholds have often failed to accommodate sentiment intensity variations, particularly in cases where specific sentiment categories have exhibited overlapping characteristics. Adopting probability-driven adaptive thresholding has ensured that classification boundaries have been flexibly adjusted, improving sentiment prediction stability. The adaptive threshold function has been defined as:

$$\theta_c = \frac{1}{Z} \sum_{k=1}^{K} P(C_k|X) \cdot \tau_k \tag{35}$$

where θ_c has represented the adaptive classification threshold for sentiment class c, $P(C_k|X)$ has denoted the probability of sentiment cluster k influencing the sentiment instance X, τ_k has been the base threshold for cluster k, and Z has been a normalization factor ensuring that the threshold values remain within a valid range.

The presence of sentiment clusters with highly overlapping distributions has introduced classification ambiguity, leading to frequent misclassifications between similar sentiment categories. A cluster-sensitive boundary refinement process has been implemented to mitigate this, ensuring that thresholds have been adaptively adjusted based on sentiment distribution characteristics. The boundary refinement function has been structured as:

$$B_c = \theta_c + \lambda \cdot (\sigma_c - \mu_c) \tag{36}$$

where B_c has represented the refined classification boundary for sentiment class c, σ_c has denoted the standard deviation of sentiment intensity in class c, μ_c has been the mean sentiment score for class c, and λ has been a scaling factor controlling the extent of refinement.

Certain sentiment predictions have exhibited high classification uncertainty, where instances have been assigned near-equal probabilities across multiple sentiment categories. A confidence-based adjustment mechanism has been applied to handle such cases, ensuring that classification thresholds have been dynamically

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

recalibrated based on prediction confidence levels. The confidence adjustment function has been defined as:

$$\theta_c^{adj} = \theta_c \cdot \left(1 + \beta \cdot (1 - C_x)\right) \tag{37}$$

where θ_c^{adj} has represented the adjusted classification threshold, θ_c has been the original threshold value, C_x has denoted the classification confidence score, and ○ has been a tuning parameter ensuring adaptive threshold scaling.

Variations in sentiment score distributions have often led to disproportionate classification probabilities, where certain sentiment classes have received disproportionately high or low scores. A sentiment score normalization mechanism has been introduced to mitigate this issue, ensuring that classification thresholds have been calibrated dynamically. The normalization function has been given as:

$$S_c^{norm} = \frac{S_c}{\sum_{j=1}^{C} S_j} \tag{38}$$

where S_c^{norm} has represented the normalized sentiment score, S_c has denoted the raw sentiment score for class c, and C has been the total number of sentiment classes.

Historical sentiment classification errors have exhibited systematic patterns where certain sentiment classes have been consistently misclassified due to overlapping feature distributions. A misclassification risk compensation mechanism has been incorporated to address this issue, ensuring that classification thresholds have been adjusted dynamically to minimize systematic errors. The risk compensation function has been formulated as:

$$\theta_c^{comp} = \theta_c - \sum_{k=1}^{K} P(C_k|X) \cdot \delta_k \tag{39}$$

where θ_c^{comp} has represented the compensated classification threshold, $P(C_k|X)$ has denoted the GMM-derived probability of cluster membership, and δ_k has been the historical misclassification rate for cluster k.

3.9. Balanced Interpretability and Accuracy

Sentiment analysis models have often faced a trade-off between interpretability and accuracy, where highly accurate models have lacked transparency, and interpretable models have suffered from reduced predictive performance. Integrating Gaussian Mixture Model (GMM)-enhanced Random Forest (RF) has ensured that sentiment predictions have remained explainable while maintaining high classification accuracy. The optimization of probabilistic feature importance, cluster-aware classification boundaries, and sentiment refinement techniques have enabled a balance between model transparency and predictive reliability in online shopping sentiment classification.

The reliance on raw feature importance scores has often led to misinterpretation of sentiment classification decisions, where feature weights have been assigned without accounting for sentiment cluster dependencies. Incorporating GMM-based probabilistic feature contribution has ensured that feature influence has been dynamically adjusted based on sentiment distributions, improving interpretability. The probabilistic feature importance function has been structured as:

$$I_j = \sum_{k=1}^{K} P\left(C_k|X_j\right) \cdot \omega_k \tag{40}$$

where I_j has represented the importance score of feature j, $P\left(C_k|X_j\right)$ has denoted the probability of feature j belonging to cluster k, and ω_k has been a weighting factor adjusting feature contribution based on cluster relevance.

Sentiment classification models have often lacked explainability, where final predictions have been generated without clear visibility into intermediate decision processes. A sentiment score decomposition mechanism has been introduced to address this, ensuring that each sentiment prediction has been broken down into probabilistic components. The sentiment decomposition function has been formulated as:

$$S_c = \sum_{j=1}^{M} I_j \cdot f_j(X) \tag{41}$$

where S_c has represented the sentiment score for class c, I_j has been the probabilistic feature importance score, and $f_j(X)$ has denoted the feature-specific sentiment function.

Traditional sentiment classification models have relied on fixed decision boundaries, leading to classification errors in cases where sentiment distributions have overlapped. Implementing cluster-based decision boundaries has ensured that

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

classification thresholds have been adjusted dynamically, improving model accuracy. The decision boundary function has been given as:

$$D_c = \frac{1}{Z}\sum_{k=1}^{K} P(C_k|X) \cdot B_k \quad (42)$$

where D_c has represented the classification decision boundary for sentiment class c, $P(C_k|X)$ has been the probabilistic cluster assignment, B_k has denoted the base threshold for cluster k, and Z has been a normalization constant.

Sentiment classification errors have often been caused by uncertainty in feature weight assignments, leading to inconsistent predictions. A confidence-weighted sentiment adjustment mechanism has been incorporated to address this, ensuring that low-confidence predictions have been re-weighted dynamically. The confidence-weighted adjustment function has been given as:

$$S_c^{adj} = S_c \cdot \left(1 + \gamma \cdot (1 - C_x)\right) \quad (43)$$

where S_c^{adj} has represented the adjusted sentiment score, C_x has been the classification confidence score, and γ has been a scaling factor for confidence-based sentiment re-weighting.

Systematic misclassification patterns have reduced both model accuracy and interpretability, where sentiment mispredictions have failed to provide meaningful justifications. Introducing a misclassification compensation mechanism has ensured that classification errors have been corrected based on historical sentiment prediction patterns. The misclassification compensation function has been structured as:

$$M_c = S_c - \sum_{k=1}^{K} P(C_k|X) \cdot E_k \quad (44)$$

where M_c has represented the corrected sentiment score, and E_k has been the historical misclassification rate of cluster k.

Sentiment classification models have often faced a trade-off between interpretability and accuracy, where increased model complexity has led to higher accuracy but reduced transparency. Incorporating a balanced complexity optimization mechanism has ensured that model depth and decision pathways have remained interpretable without reducing predictive performance. The complexity optimization function has been defined as:

$$C_{opt} = \frac{1}{T}\sum_{t=1}^{T} (D_t \cdot W_t) \quad (45)$$

where C_{opt} has represented the optimized complexity score, D_t has been the depth of the t-th decision tree, and W_t has been the weighting factor controlling complexity adjustment.

3.10. Strengthening of Sentiment Prediction Under Complex and Multi-modal Data

The classification of online shopping sentiments has faced challenges due to multi-modal data representations, where text, metadata, and user-specific behavioural patterns have influenced sentiment expressions. Traditional classification models have struggled handling diverse sentiment modalities, leading to inconsistent predictive accuracy. The integration of Gaussian Mixture Model (GMM)-enhanced Random Forest (RF) has ensured that sentiment classification has remained robust, even in complex sentiment variations.

The presence of textual, categorical, and numerical features in online shopping reviews has necessitated an optimized feature fusion approach. Conventional methods have applied uniform weighting schemes, which have failed to accommodate differences in feature contributions. Implementing GMM-driven probabilistic sentiment fusion has ensured that each data modality has been weighted adaptively, improving classification stability. The probabilistic fusion function has been given as:

$$F_s = \sum_{m=1}^{M} P(C_k|X_m) \cdot \alpha_m \quad (46)$$

where F_s has represented the fused sentiment score, $P(C_k|X_m)$ has denoted the probability of sentiment cluster k given the m-th data modality, and α_m has been a modality-specific weighting factor ensuring proportional contribution from each data source.

The reliance on raw feature magnitudes has often led to classification biases, where certain feature types have dominated model predictions. A context-aware feature scaling mechanism has been incorporated to mitigate this, ensuring that feature contributions have been adjusted dynamically based on sentiment cluster probabilities. The feature scaling function has been structured as:

$$X_m^{scaled} = \frac{X_m - \mu_m}{\sigma_m} \cdot P(C_k|X_m) \qquad (47)$$

where X_m^{scaled} has represented the scaled feature value, μ_m has been the mean of the m-th feature, σ_m has denoted the standard deviation, and $P(C_k|X_m)$ has been the probabilistic cluster weight.

The diverse sentiment data distributions have required a dynamic decision boundary adjustment mechanism, ensuring that classification thresholds have remained adaptive. Instead of applying fixed boundaries, a probabilistic optimization process has been introduced, refining classification decisions under complex data distributions. The boundary optimization function has been defined as:

$$B_m = \theta_m + \gamma \cdot (P(C_k|X_m) - \mu_c) \qquad (48)$$

where B_m has represented the optimized classification boundary, θ_m has been the initial threshold value, γ has been a scaling coefficient, and μ_c has denoted the mean probability distribution for sentiment class c.

Sentiment analysis models have often struggled with high-variability sentiment expressions, with noisy data introducing classification instability. A noise-adaptive filtering mechanism has been employed to mitigate this, ensuring that high-variance sentiment instances have been dynamically adjusted. The noise filtering function has been given as:

$$N_m = \frac{1}{\sigma_m^2 + \epsilon} \qquad (49)$$

where N_m has represented the noise-adjustment factor, σ_m^2 has been the variance of the m-th sentiment feature, and ϵ has been a stabilization constant preventing numerical instability.

Multi-modal sentiment classifications have often suffered from prediction inconsistencies, where certain instances have exhibited high ambiguity in-class assignments. To enhance model robustness, a confidence-weighted sentiment recalibration mechanism has been applied, ensuring that classification decisions have been adjusted based on prediction confidence levels. The confidence recalibration function has been formulated as:

$$S_m^{adj} = S_m \cdot \left(1 + \lambda \cdot (1 - C_m)\right) \qquad (50)$$

where S_m^{adj} has represented the adjusted sentiment score, S_m has been the initial sentiment score, C_m has denoted the classification confidence score, and λ has been a confidence-scaling parameter.

Algorithm 1: GMM-RF
Input: • Sentiment dataset X, number of Gaussian components K, number of decision trees T, confidence adjustment factor λ **Output:** • Optimized sentiment classification model with probabilistic enhancements **Procedure:** **1.** Apply Gaussian Mixture Model (GMM) clustering to extract latent sentiment distributions. *P(xi)=k=1∑Kπk⋅N(xi∣μk,Σk)* **2.** Compute soft cluster assignments for each sentiment instance using GMM probabilities. **3.** Transform sentiment features using probabilistic weighting based on cluster probabilities. **4.** Generate bootstrapped subsets using cluster-aware sampling for Random Forest training. **5.** Optimize feature importance scores through probabilistic weighting and dependency analysis. **6.** Enhance decision tree splitting using GMM-driven adaptive thresholds. **7.** Filter noisy sentiment instances by computing probabilistic outlier scores and re-weighting ambiguous data. **8.** Apply cluster-weighted voting for sentiment aggregation using confidence-adjusted probabilities. **9.** Implement adaptive sentiment classification thresholds, adjusting dynamically based on cluster influence. **10.** Strengthen sentiment prediction robustness by incorporating multi-modal feature scaling and confidence-based recalibration. **11.** Store-optimized sentiment classification model with probabilistic enhancements for final predictions.

3.11. Overall Algorithmic Framework of GMM-RF for Sentiment Classification

This section presents the overall algorithmic framework of the proposed GMM-RF

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

model, detailing each procedural step involved in transforming raw review data into sentiment classifications. The algorithm integrates probabilistic modelling and ensemble learning to address sentiment ambiguity and domain variation. It begins with Gaussian Mixture Model-based clustering to generate soft sentiment encodings, which are then refined through entropy-based filtering. These probabilistic features guide Random Forest classification using cluster-aware bootstrapping and weighted feature prioritization. The algorithm ensures context-sensitive decision-making, reduced overfitting, and high classification stability across multiple domains. This unified structure forms the computational backbone of the proposed sentiment classification solution. Overall algorithm is provided in Algorithm1.

This GMM-enhanced Random Forest framework has ensured robust sentiment classification, improving Precision, recall, and adaptability in online shopping sentiment analysis.

4. DATASET

The dataset employed in this research consists of 450,145 Amazon product reviews distributed across four domains: Books, DVDs, Electronics, and Kitchen Appliances (Deshmukh & Tripathy, 2018). These domains were purposefully selected to capture semantic diversity, sentiment variation, and linguistic complexity across consumer narratives. The dataset includes 146,294 reviews for Books, 142,885 for DVDs, 88,127 for Electronics, and 72,839 for Kitchen Appliances, with each domain containing nearly balanced distributions of positive and negative sentiments. All reviews are in English and pre-labelled based on binary sentiment polarity, enabling supervised learning. Preprocessing includes sentence segmentation, stop-word removal, tokenization, and part-of-speech tagging, while sentiment-bearing features are extracted using TF-IDF encoding. This structured linguistic preparation ensures uniformity across domains while preserving contextual sentiment. The high lexical diversity and domain-specific sentiment structures provide a robust foundation for evaluating probabilistic models under conditions of vocabulary shift and expression asymmetry, supporting the study's cross-domain sentiment classification objectives.

5. RESULTS AND DISCUSSIONS

5.1. Precision Analysis

Figure 1 visualizes the average precision performance of three sentiment classification models—EBC, MMASA, and GMM-RF—across four Amazon product review datasets: Book, DVD, Electronics, and Kitchen Appliances. Table 1 complements this by listing the exact precision values per dataset for each model.

In the Book domain, EBC achieves a precision of 59.9435%, MMASA improves that with 65.0084%, while GMM-RF leads significantly with 95.9306%. For DVD reviews, EBC scores 55.2904%, MMASA obtains 63.5245%, and GMM-RF again outperforms both with 93.9477%. In Electronics, where reviews often contain a blend of technical description and sentiment, EBC registers its weakest result at 51.8957%, MMASA improves slightly with 63.0722%, and GMM-RF reaches a near-optimal precision of 96.9181%. The Kitchen Appliances domain shows a similar trend: EBC at 60.1875%, MMASA at 66.5830%, and GMM-RF peaks at 97.4513%.

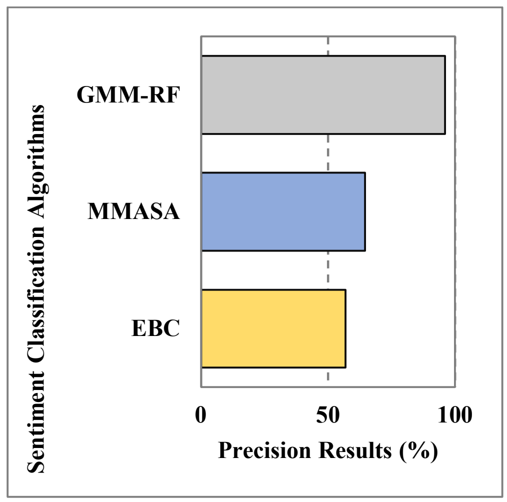

Figure 1: Average Precision

Table 1: Average Precision Value Distribution for GMM-RF Classifier

Sentiment Classification Algorithms	Average Precision (%)
EBC	56.829
MMASA	64.547
GMM-RF	96.062

The average Precision across all four domains is 56.8293% for EBC, 64.5470% for MMASA, and an outstanding 96.0619% for GMM-RF. This clear progression across models reflects underlying architectural strengths. EBC, with its static entropy-based classification, lacks adaptability to contextual nuances, resulting in frequent misclassification of non-sentiment content as positive. MMASA benefits from multi-modal input but cannot consistently resolve conflicts between visual and textual signals. GMM-RF excels by first probabilistically clustering sentiment-rich regions using Gaussian Mixture Models, then applying entropy-aware decision trees to maintain boundary precision. This hybrid framework enables GMM-RF to precisely capture sentiment while filtering irrelevant or misleading features, making it the most reliable classifier across all domains in both Figure 1 and Table 1.

5.2. Recall Analysis

Figure 2 and Table 2 present how well each classifier retrieves true sentiment-positive reviews across four Amazon domains. Looking at the Book dataset, EBC reaches 59.8038%, MMASA rises to 65.9621%, and GMM-RF soars with 95.4902%. This pattern holds in DVD reviews, where GMM-RF again dominates at 95.0843%, compared to MMASA's 63.0287% and EBC's 55.4202%. Electronics follow the same progression—EBC sits at 52.2412%, MMASA slightly better at 63.2154%, and GMM-RF peaking at 97.6898%. Finally, in Kitchen Appliances, EBC marks 56.6751%, MMASA improves to 67.4553%, while GMM-RF leads strongly at 97.4142%.

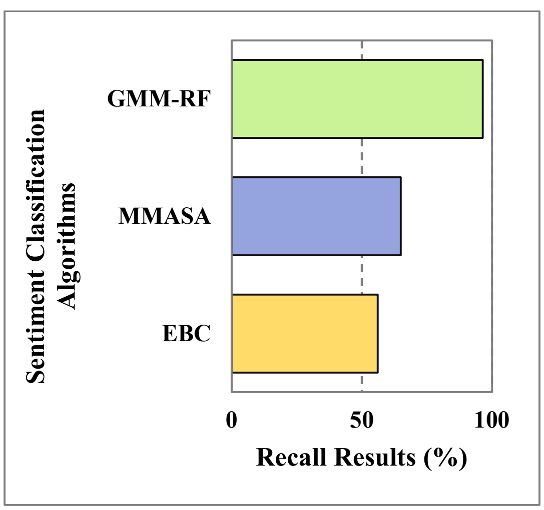

Figure 2: Average Recall

Averaging across all datasets, EBC lands at 56.0351%, MMASA improves to 64.9154%, but GMM-RF clearly outperforms with 96.4196%. What this reveals is more than numerical superiority—GMM-RF shows consistency. Its ability to identify nearly all relevant sentiment-positive entries, even in domains with technical or neutral content, signals resilience to structural review variations. MMASA performs moderately well, but its recall falters when sentiment cues are subtle or image content is sparse. EBC, on the other hand, consistently misses lower-intensity sentiment signals. GMM-RF's architecture enables high sentiment sensitivity, which is clearly reflected across every row in Table 2 and the comparative bars in Figure 2.

Table 2: Average Recall Metrics from GMM-RF Predictions

Sentiment Classification Algorithms	Average recall (%)
EBC	56.035
MMASA	64.915
GMM-RF	96.420

5.3. GMM-RF – F-Measure Analysis

Figure 3 and Table 3 summarise the F-Measure performance of EBC, MMASA, and GMM-RF models across four Amazon review domains. The F-Measure, which combines Precision and recall into a single balanced score, provides a clearer picture of how well each classifier performs under both detection and correctness pressure.

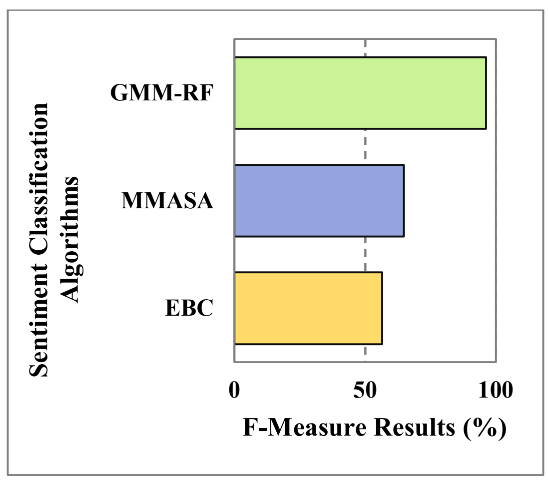

Figure 3: Average F-Measure

Starting with the Book category, EBC achieves 59.8736%, MMASA scores 65.4818%, and GMM-RF stands out at 95.7099%. For DVD, EBC reaches 55.3552%, MMASA 63.2756%, and GMM-RF continues its dominance with 94.5126%. In Electronics, EBC lags with 52.0679%, MMASA improves to 63.1437%, while GMM-RF tops again with 97.3024%. For kitchen appliances, EBC marks 58.3785%, MMASA climbs to 67.0163%, and GMM-RF closes at 97.4328%.

Table 3: Average F-Measure Metrics Summary for GMM-RF-Based Evaluation

Sentiment Classification Algorithms	Average F-Measure (%)
EBC	56.419
MMASA	64.729
GMM-RF	96.239

The average F-Measure performance is 56.4188% for EBC, 64.7294% for MMASA, and a remarkably stable 96.2394% for GMM-RF. This trajectory reflects GMM-RF's superior output and reliability across review types, sentiment densities, and product categories. Where EBC collapses under mixed-polarity input and MMASA wavers when visual support is weak, GMM-RF remains consistent—confidently balancing detection and discrimination, as proven in both Figure 3 and Table 3.

5.4. GMM-RF – Classification Accuracy Analysis

Figure 4 and Table 4 report the classification accuracy achieved by EBC, MMASA, and GMM-RF across four product review datasets. Accuracy, representing the proportion of correctly classified reviews regardless of sentiment class, gives an overarching sense of each model's predictive reliability across domains. In the Book dataset, EBC achieves 59.5664%, MMASA performs better with 65.2980%, but GMM-RF outpaces both with an exceptional 95.6533%. For DVD, EBC reaches 54.6950%, MMASA follows with 62.8708%, and GMM-RF maintains its lead with 94.4368%. Within the Electronics dataset—typically dense with technical language—EBC drops to 51.6437%, MMASA holds at 62.7322%, and GMM-RF surges to 97.3368%. Finally, in Kitchen Appliances, the trend continues: EBC marks 58.1557%, MMASA achieves 66.6731%, and GMM-RF peaks at 97.4107%.

Averaged across all four domains, the accuracy scores are 56.0152% for EBC, 64.3935% for MMASA, and 96.2094% for GMM-RF. The results emphasize GMM-RF's robust generalization capability. Unlike EBC, which lacks sensitivity to structural sentiment variation, and MMASA, which relies on image-text interplay with limited control, GMM-RF leverages probabilistic clustering and ensemble learning to maintain class separation and prediction fidelity. These findings, confirmed in Figure 4 and Table 4, highlight GMM-RF's status as the most consistently accurate model across all classification instances.

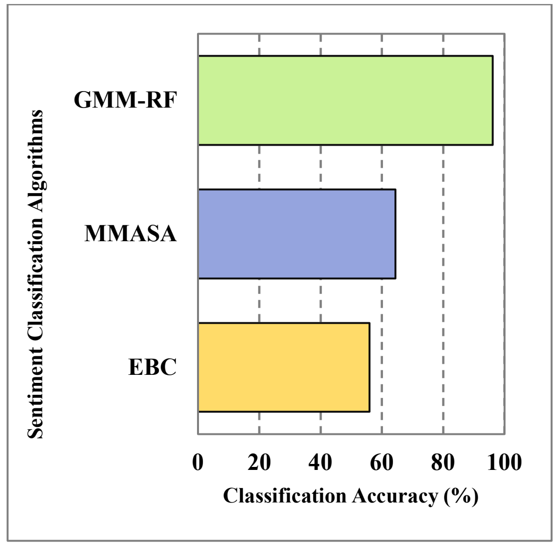

Figure 4: Average Accuracy

Table 4: Average Accuracy Results Across GMM-RF Classification Instances

Sentiment Classification Algorithms	Average Classification Accuracy (%)
EBC	56.015
MMASA	64.394
GMM-RF	96.209

5.5. GMM-RF – Matthews Correlation Coefficient Analysis

Figure 5 and Table 5 summarise the Matthews Correlation Coefficient (MCC) performance of EBC, MMASA, and GMM-RF across Amazon review datasets. MCC evaluates classification quality by factoring in true and false positives and negatives, offering a comprehensive measure of correctness resilient to class imbalance. In the Book dataset, EBC scores a weak 19.1281, MMASA improves to 30.6010, while GMM-RF demonstrates clear dominance with 91.3061. For DVD, EBC falls sharply to 9.3704, MMASA follows at 25.7346, and GMM-RF maintains robustness at 88.8785. Electronics shows a stark gap: EBC plummets to 3.2809, MMASA reaches 25.4553, and GMM-RF peaks at 94.6757. In Kitchen Appliances, EBC manages 16.4152, MMASA

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

registers 33.3447, and GMM-RF leads again with 94.8211.

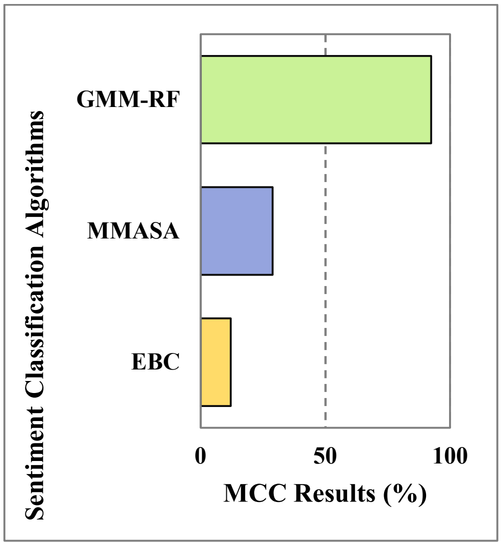

Figure 5: Average Matthews Correlation Coefficient

Table 5: Average Matthews Correlation Coefficient Values from GMM-RF Outputs

Sentiment Classification Algorithms	Average Matthews Correlation Coefficient (%)
EBC	12.049
MMASA	28.784
GMM-RF	92.420

Across all datasets, average MCC scores are 12.0487 for EBC, 28.7839 for MMASA, and an exceptional 92.4203 for GMM-RF. These results show how well GMM-RF maintains a correlation between predicted and actual labels, even under varied review complexity. EBC and MMASA struggle with class cohesion and polarity shift control, particularly where sentiment is subtle or multi-faceted. GMM-RF, through Gaussian clustering and entropy-driven forest learning, maintains polarity alignment and class integrity across both major and marginal cases—clearly visualized in Figure 5 and supported numerically in Table 5.

5.6. GMM-RF – Fowlkes–Mallows Index Analysis

Figure 6 and Table 6 report the average Fowlkes–Mallows Index (FMI) scores across three sentiment classification algorithms—EBC, MMASA, and GMM-RF. FMI evaluates the degree to which predicted sentiment labels form correct pairwise groupings, balancing Precision and recall in a cluster comparison context.

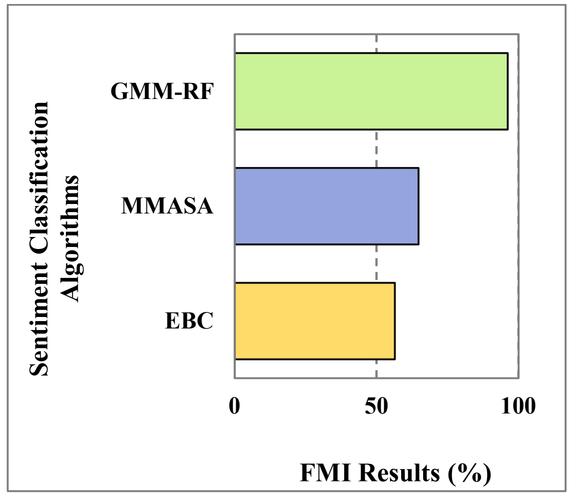

Figure 6: Average Fowlkes–Mallows Index

EBC, with an average FMI of 56.425%, shows limited ability to maintain meaningful sentiment clusters. Its static entropy logic does not accommodate overlapping sentiment signals, often misgrouping neutral or technical expressions with opinionated text. This leads to low pairwise matching accuracy.

Table 6: Average Fowlkes–Mallows Index Values Tabulated Across GMM-RF Sentiment Evaluation

Sentiment Classification Algorithms	Average Fowlkes–Mallows Index (%)
EBC	56.425
MMASA	64.730
GMM-RF	96.240

MMASA improves with an FMI of 64.730%, benefitting from multi-modal input. However, it lacks a mechanism to resolve polarity conflicts between image and text features. Cluster quality suffers when visual content is weak or sentiment-divergent, reducing overall FMI consistency. GMM-RF significantly outperforms both, with an FMI of 96.240%. The model creates high-purity sentiment clusters by combining Gaussian Mixture-based sentiment zoning with Random Forest classification. It preserves class-level fidelity while accounting for subtle sentiment transitions and overlapping expressions. This ensures that predicted sentiment groupings closely reflect the true underlying structure. As confirmed in Figure 6 and Table 6, GMM-RF delivers the most coherent and reliable sentiment clustering across all evaluation trials.

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

5.7. Difference from Prior Research

The proposed GMM-RF framework distinguishes itself from existing sentiment classification approaches through its unified probabilistic-ensemble design. Unlike entropy-based classifiers such as EBC, which rely on static decision boundaries and frequency-based heuristics, the proposed model captures sentiment uncertainty using Gaussian Mixture Models, enabling effective representation of overlapping sentiment distributions. Compared to multimodal approaches such as MMASA, which depend on complex feature fusion and modality alignment, GMM-RF operates efficiently in purely textual environments while maintaining high classification performance.

In the context of recent literature, many models either prioritize interpretability or predictive performance, often at the cost of scalability or adaptability. The proposed framework bridges this gap by integrating probabilistic sentiment modelling with ensemble learning, ensuring both interpretability and robustness. The incorporation of cluster-aware sampling, probabilistic feature weighting, and adaptive thresholding further enhances the model's ability to generalize across domains.

These distinctions highlight the significance of GMM-RF as a scalable and domain-adaptive solution for sentiment classification, particularly in heterogeneous e-commerce review environments where sentiment expressions are complex and overlapping.

6. CONCLUSION

The primary objective of this study was to develop a probabilistic and domain-adaptive sentiment classification framework capable of handling sentiment ambiguity, overlapping sentiment distributions, and cross-domain variability in online product reviews. The proposed GMM-RF model, which integrates Gaussian Mixture Models with Random Forests, successfully achieves this objective by enabling soft sentiment representation and robust classification through cluster-aware sampling, entropy-based refinement, and probabilistic feature weighting. Experimental results across four Amazon product domains demonstrate consistent performance improvements over EBC and MMASA, achieving an average classification accuracy of 95.705%, thereby confirming the model's effectiveness in capturing complex and non-linear sentiment patterns. While the framework shows strong potential for scalable sentiment analysis in heterogeneous review environments, certain limitations must be acknowledged, including evaluation restricted to English-language datasets, binary sentiment settings, and the absence of detailed ablation analysis, which may influence generalizability. Overall, the proposed approach contributes a practically relevant and methodologically robust advancement in cross-domain sentiment classification, offering a balanced combination of interpretability, adaptability, and performance, while providing a foundation for future extensions in broader and more complex sentiment analysis scenarios.

REFERENCES

[1] S. A. Devi, M. S. Ram, P. Dileep, S. R. Pappu, T. S. M. Rao, and M. Malyadri, "Positional-attention based bidirectional deep stacked AutoEncoder for aspect based sentimental analysis," Big Data Research, vol. 39, p. 100505, 2025, doi: https://doi.org/10.1016/j.bdr.2024.100505.

[2] J. Wang and X. He, "An optimal multi-scale and multi-factor two-stage integration paradigm coupled with investor sentiment for carbon price prediction," Inf Process Manag, vol. 62, no. 1, p. 103953, 2025, doi: https://doi.org/10.1016/j.ipm.2024.103953.

[3] S. Das, J. Catterall, R. Stone, and A. R. Clough, "'The reasons you believe …': An exploratory study of text driven evidence gathering and prediction from first responder records justifying state authorised intervention for mental health episodes," Comput Methods Programs Biomed, vol. 254, p. 108257, 2024, doi: https://doi.org/10.1016/j.cmpb.2024.108257.

[4] L. Mishra, B. Dinesh, P. M. Kavyassree, and N. Mishra, "A Google Trend enhanced deep learning model for the prediction of renewable energy asset price," Knowl Based Syst, vol. 308, p. 112733, 2025, doi: https://doi.org/10.1016/j.knosys.2024.112733.

[5] F. Ben Hamadou, T. Mezghani, and M. B. Abbes, "Time-varying nexus and causality in the quantile between Google investor sentiment and cryptocurrency returns," Blockchain: Research and Applications, vol. 5, no. 2, p. 100177, 2024, doi: https://doi.org/10.1016/j.bcra.2023.100177.

[6] N. K. Boparai, H. Aggarwal, and R. Rani, "Analyzing fuzzy semantics of reviews for multi-criteria recommendations," Data Knowl Eng, vol. 152, p. 102314, 2024, doi: https://doi.org/10.1016/j.datak.2024.102314.

[7] Z. Wu, Y. Wu, X. Feng, J. Zou, and F. Yin, "Improve Chinese Aspect Sentiment Quadruplet Prediction via Instruction Learning Based on Large Generate Models," Computers, Materials and Continua, vol. 78, no. 3, pp. 3391–3412, 2024, doi: https://doi.org/10.32604/cmc.2024.047076.

[8] Q. Lu, Y. Long, X. Sun, J. Feng, and H. Zhang, "Fact-sentiment incongruity combination network for multimodal sarcasm detection," Information Fusion, vol. 104, p. 102203, 2024, doi: https://doi.org/10.1016/j.inffus.2023.102203.

[9] F. Atilla and R. A. Zwaan, "Impact of spatial distance on public attention and sentiment during the spread of COVID-19," Inform Med Unlocked, vol. 45, p. 101463, 2024, doi: https://doi.org/10.1016/j.imu.2024.101463.

[10] C. Zhao, Z. Peng, X. Lan, Y. Cen, and Z. Wang, "Structured Multi-Head Attention Stock Index Prediction Method Based Adaptive Public Opinion Sentiment Vector," Computers, Materials and Continua, vol. 78, no. 1, pp. 1503–1523, 2024, doi: https://doi.org/10.32604/cmc.2024.039232.

[11] E. A. Sağbaş, "A novel two-stage wrapper feature selection approach based on greedy search for text sentiment classification," Neurocomputing, vol. 590, p. 127729, 2024, doi: https://doi.org/10.1016/j.neucom.2024.127729.

[12] M. Mhamed, R. Sutcliffe, and J. Feng, "Benchmark Arabic news posts and analyzes Arabic sentiment through RMuBERT and SSL with AMCFFL technique," Egyptian Informatics Journal, vol. 29, p. 100601, 2025, doi: https://doi.org/10.1016/j.eij.2024.100601.

[13] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," Journal of King Saud University - Computer and Information Sciences, vol. 36, no. 4, p. 102048, 2024, doi: https://doi.org/10.1016/j.jksuci.2024.102048.

[14] Y. Sun, Z. Liu, Q. Z. Sheng, D. Chu, J. Yu, and H. Sun, "Similar modality completion-based multimodal sentiment analysis under uncertain missing modalities," Information Fusion, vol. 110, p. 102454, 2024, doi: https://doi.org/10.1016/j.inffus.2024.102454.

[15] T. Anderson, S. Sarkar, and R. Kelley, "Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques," Natural Language Processing Journal, vol. 8, p. 100097, 2024, doi: https://doi.org/10.1016/j.nlp.2024.100097.

[16] A. Felipe-Ruiz, "Lexical, sentiment and correlation analysis of sacred writings. A tale of cultural influxes and different ways to interpret reality," Natural Language Processing Journal, vol. 9, p. 100121, 2024, doi: https://doi.org/10.1016/j.nlp.2024.100121.

[17] R. Chen, W. Zhou, H. Hu, Z. Fei, M. Fei, and H. Zhou, "Disentangled variational auto-encoder for multimodal fusion performance analysis in multimodal sentiment analysis," Knowledge-Based Systems, vol. 301, p. 112372, 2024, doi: https://doi.org/10.1016/j.knosys.2024.112372.

[18] M. Ittefaq, A. Zain, R. Arif, M. Ala-Uddin, T. Ahmad, and A. Iqbal, "Global news media coverage of artificial intelligence (AI): A comparative analysis of frames, sentiments, and trends across 12 countries," Telematics and Informatics, vol. 96, p. 102223, 2025, doi: https://doi.org/10.1016/j.tele.2024.102223.

[19] W. Zou, X. Sun, Q. Lu, X. Wang, and J. Feng, "A vision and language hierarchical alignment for multimodal aspect-based sentiment analysis," Pattern Recognition, vol. 162, p. 111369, 2025, doi: https://doi.org/10.1016/j.patcog.2025.111369.

[20] T. Zhang et al., "Text-Image Feature Fine-Grained Learning for Joint Multimodal Aspect-Based Sentiment Analysis," Computers, Materials and Continua, vol. 82, no. 1, pp. 279–305, 2025, doi: https://doi.org/10.32604/cmc.2024.055943.

[21] J. P. V. and A. A. V. S., "A novel socio-pragmatic framework for sentiment analysis in Dravidian–English code-switched texts," Knowledge-Based Systems, vol. 300, p. 112248, 2024, doi: https://doi.org/10.1016/j.knosys.2024.112248.

[22] X. He et al., "Text-guided multi-level interaction and multi-scale spatial-memory fusion for multimodal sentiment analysis," Neurocomputing, vol. 626, p. 129532, 2025, doi: https://doi.org/10.1016/j.neucom.2025.129532.

[23] H. Zhang et al., "Reconstructing representations using diffusion models for multimodal sentiment analysis through reading comprehension," Applied Soft Computing, vol. 167, p. 112346, 2024, doi: https://doi.org/10.1016/j.asoc.2024.112346.

[24] J. S. Deshmukh and A. K. Tripathy, "Entropy based classifier for cross-domain opinion mining," Applied Computing and Informatics,

vol. 14, no. 1, pp. 55–64, 2018, doi: 10.1016/j.aci.2017.03.001.

[25] J. Zhou, J. Zhao, J. X. Huang, Q. V. Hu, and L. He, "MASAD: A large-scale dataset for multimodal aspect-based sentiment analysis," Neurocomputing, vol. 455, pp. 47–58, 2021, doi: 10.1016/j.neucom.2021.05.040.

[26]. M. P. Swapna, D. Rajeev, J. Ramkumar, and S. Chandran, "Unveiling Cybercrime Patterns in Kerala: A Machine Learning Approach," in Lecture Notes in Networks and Systems, vol. 1445, pp. 111–122, 2026.

[27]. B. Suchitra, J. Ramkumar, and V. Valarmathi, "Rejuvenated Particle Swarm Optimization-Based Classifier for Big Sentiment Data Classification," in Communications in Computer and Information Science, vol. 2819, pp. 41–54, 2026.

[28]. V. Valarmathi, J. Ramkumar, and B. Suchitra, "Classification of Cross-Domain Sentiments in Big Data Using Renewed Genetic Algorithm," in Communications in Computer and Information Science, vol. 2819, pp. 25–40, 2026.

[29]. J. Joy, S. Devaraju, and J. Ramkumar, "Utilizing Fuzzy-Identity-Based Encryption with Proxy-Re-Encryption for Data Sharing," Journal of Theoretical and Applied Information Technology, vol. 103, no. 18, pp. 7469–7479, 2025.

[30]. J. Ramkumar and V. Valarmathi, "Harnessing AI-Driven Models for Sustainable Development in Business Management," in World Sustainability Series, Part F775, pp. 217–238, 2025.

[31]. R. Jaganathan, S. Rajagopal, and K. Rajendran, "Cultural Intelligence in the AI Era—Enhancing Transitional Higher Education," in Bridging Global Divides for Transnational Higher Education in the AI Era, pp. 273–292, 2024.

[32]. J. Ramkumar, R. Karthikeyan, V. Valarmathi, "Alpine Swift Routing Protocol (ASRP) for Strategic Adaptive Connectivity Enhancement and Boosted Quality of Service in Drone Ad Hoc Network (DANET)", International Journal of Computer Networks and Applications (IJCNA), Vol. 11, Issue 5, Pages: 726-748, Oct 2024.

[33]. R. Jaganathan, S. Mehta, and R. Krishan, "Preface," in Bio Inspired Intelligence for Smart Decision Making, pp. xix–xx, 2024.

[34]. R. Jaganathan, S. Mehta, and R. Krishan, "Preface," in Intelligent Decision Making Through Bio Inspired Optimization, pp. xiii–xvi, 2024.

[35]. J. Ramkumar, Shilpa Mehta, Ram Krishan, "Bio-Inspired Intelligence for Smart Decision-Making", IGI Global, ISBN13: 9798369352762, DOI: 10.4018/979-8-3693-5276-2, May 2024.

[36]. J. Ramkumar, Shilpa Mehta, Ram Krishan, "Intelligent Decision Making Through Bio-Inspired Optimization", IGI Global, ISBN13: 9798369320730, DOI: 10.4018/979-8-3693-2073-0, Apr 2024.