International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(22s) 2026 pp. 153-178
https://doi.org/10.70917/ijcisim-2026-5423
Article

Hybrid LaBSE Semantic and Handcrafted Feature Fusion with Machine Learning for Fake Review Detection in Roman Marathi Code-Mixed Text

Swapnil S. Nehar[1], Ranjit R. Keole[2], Pravin P. Karde[3]

[1]Department of Computer Science and Engineering, HVPM COET, Amravati (M.S), India. Email: swapnil.nehar@gmail.com
[2]Department of Information Technology, HVPM COET, Amravati (M.S), India. Email: ranjitkeole@gmail.com
[3]Department of Information Technology, Government Polytechnic, Amravati (M.S), India. Email: ppkarde@gmail.com

Abstract: Fake review detection in low-resource and code-mixed languages remains challenging due to informal writing styles, transliterated regional expressions, linguistic variability, and the limited availability of annotated datasets. This paper presents a hybrid LaBSE semantic and handcrafted feature fusion approach with machine learning for fake review detection in Roman Marathi code-mixed text. A real-time dataset comprising 2,287 Roman Marathi reviews collected from multiple online platforms is utilized to evaluate the proposed approach. The framework integrates opinion-mining features with multilingual semantic representations generated using Language-agnostic BERT Sentence Embedding (LaBSE) and handcrafted linguistic, behavioural, contextual, temporal, and metadata features to construct a comprehensive hybrid feature representation. The dataset is balanced using random oversampling and subsequently partitioned into training and testing subsets using an 80:20 ratio. Four machine learning classifiers, namely Random Forest, XGBoost, Support Vector Machine, and K-Nearest Neighbour, are evaluated using accuracy, precision, recall, and F1-score. Experimental results demonstrate that XGBoost achieves the best performance with 94.60% accuracy, 95.63% precision, 93.47% recall, and 94.54% F1-score, outperforming the other evaluated classifiers. The findings demonstrate the effectiveness of combining multilingual semantic information with explicit linguistic and contextual characteristics for identifying deceptive reviews in Roman Marathi code-mixed environments and establish an initial benchmark for fake review detection in this low-resource setting.

Keywords: Fake Review Detection, Roman Marathi, Code-Mixed Text, LaBSE, Semantic Embeddings, Handcrafted Features, Feature Fusion, Machine Learning, XGBoost

1. INTRODUCTION

Online reviews have become an important source of information for consumers when evaluating products, services, and sellers on digital platforms. However, the increasing influence of user-generated reviews has also encouraged the proliferation of deceptive or fake reviews intended to manipulate consumer perceptions and purchasing decisions. Recent studies indicate that fake review detection has evolved from conventional linguistic and statistical approaches toward deep learning techniques capable of learning complex textual representations [1]. Graph-based approaches have also emerged to capture relationships among reviewers, products, and review activities, although their effectiveness depends considerably on the availability and quality of relational information [2].

Existing research has investigated fake review detection using textual, behavioural, contextual, and reviewer-related information. Comprehensive studies have highlighted the use of machine learning, natural language processing (NLP), sentiment analysis, and deep learning for distinguishing genuine reviews from deceptive content, while also identifying challenges associated with feature selection, dataset characteristics, and model generalization [3]. Recent

state-of-the-art analyses further indicate that the effectiveness of fake review detection depends not only on the selected classification technique but also on the representation of review content and associated behavioural information [4]. Despite considerable progress, issues such as linguistic diversity, evolving deception strategies, limited annotated datasets, and the adaptability of detection models across different domains and platforms continue to present important research challenges [5].

Machine learning remains one of the widely adopted approaches for fake review identification because of its capability to learn discriminative patterns from textual and metadata-based features. Different supervised learning algorithms have been investigated for this purpose, demonstrating varying performance depending on feature representation and dataset characteristics [6]. Comparative studies of fake review detection in e-commerce environments have similarly demonstrated the applicability of machine learning techniques for identifying deceptive review patterns [7]. In parallel, systematic investigations of machine learning and deep learning strategies have shown an increasing transition from manually engineered representations toward semantic and context-aware representations capable of capturing complex patterns within review text [8]. Nevertheless, systematic reviews of machine learning approaches have emphasized that model performance remains closely associated with appropriate preprocessing, informative feature engineering, classifier selection, and the characteristics of the underlying review dataset [9].

Deceptive opinion spam detection presents additional challenges because fraudulent reviews may intentionally imitate the linguistic characteristics of genuine user opinions. Literature on deceptive opinion spam has therefore emphasized the importance of combining textual characteristics, sentiment information, behavioural indicators, and advanced learning techniques instead of relying on a single category of features [10]. These challenges become more significant for low-resource and code-mixed languages, where informal transliteration, spelling variations, regional expressions, English vocabulary, and platform-specific terminology can reduce the effectiveness of conventional monolingual text representations. The problem is particularly relevant to Roman Marathi, in which Marathi expressions are represented using the Roman alphabet and are frequently intermixed with English words and product-related expressions.

To address these challenges, this paper proposes a hybrid LaBSE semantic and handcrafted feature fusion approach with machine learning for fake review detection in Roman Marathi code-mixed text. A real-time Roman Marathi fake review dataset containing reviews acquired from multiple online platforms is considered. The proposed framework combines opinion-mining information with multilingual semantic embeddings generated using Language-agnostic BERT Sentence Embedding (LaBSE) and handcrafted linguistic, behavioural, contextual, temporal, and metadata features. The resulting hybrid representation is subsequently evaluated using Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) classifiers. This integration is intended to capture both the implicit semantic characteristics of Roman Marathi–English code-mixed review text and explicit indicators associated with reviewer behaviour and review authenticity.

The main contributions of this work are summarized as follows:

1. A Roman Marathi code-mixed fake review detection framework is developed for addressing deceptive review identification in a comparatively low-resource linguistic environment.

2. A hybrid feature representation is constructed by integrating LaBSE-based multilingual semantic embeddings, opinion-mining information, and handcrafted linguistic, behavioural, contextual, temporal, and metadata features.

3. A real-time, multi-platform Roman Marathi fake review dataset is utilized to investigate naturally occurring code-mixed and transliterated review characteristics.

4. Multiple machine learning classifiers, including RF, XGBoost, SVM, and KNN, are comparatively evaluated using accuracy, precision, recall, F1-score, and confusion-matrix-based analysis to determine the effectiveness of the proposed hybrid representation.

The remainder of this paper is organized as follows. Section II presents the related work on machine learning, deep learning, opinion mining, semantic representations, and hybrid feature-based fake review detection. Section III describes the proposed methodology, including dataset preparation, preprocessing, opinion mining, LaBSE-based semantic feature extraction, handcrafted feature engineering, feature fusion, dataset balancing, and machine learning-based classification. Section IV presents the experimental setup, performance evaluation, experimental results, comparative analysis, and discussion of the proposed framework using the Roman Marathi code-mixed fake review dataset. Finally, Section V concludes the paper by summarizing the major findings and outlining possible directions for future research.

2. RELATED WORK

Fake review detection has been widely studied using machine learning, NLP, sentiment analysis, deep learning, transformer models, and behavioural feature analysis. Recent studies increasingly combine semantic, linguistic, contextual, and reviewer-related information to improve the distinction between genuine and deceptive reviews.

A hybrid neural-network framework using preprocessing, tokenization, CNN-based feature extraction, and neural classification showed promising training performance but comparatively limited generalization [11]. Stack ensemble learning with BoW and TF-IDF features achieved nearly 89% accuracy and outperformed several individual machine learning classifiers [12]. Psychological and linguistic indicators derived using NLP, sentiment analysis, and LIWC were also found useful for fake review identification [13]. Comparative evaluation of LSTM, CNN, RF, NB, LR, and SVM further showed that deep sequential models can effectively capture deceptive review patterns [14]. CNNs combined with word embeddings, emotion mining, and similarity analysis achieved more than 96% accuracy for fraudulent Amazon review detection [15]. A BERT-based hybrid model integrating contextual and sentiment information reported an F1-score of 0.9466 [16]. TF-IDF and BoW representations with conventional classifiers also remained effective, with Linear SVC achieving 87.9% accuracy [17]. The combination of LSTM, BERT embeddings, sentiment, and metadata further demonstrated the benefit of integrating semantic and contextual information [18].

Transformer-based approaches have shown strong capability in learning contextual representations. An optimized DeBERTa framework achieved approximately 98% accuracy across multiple datasets [19]. Weighted fusion of BERT-based contextual features and reviewer behaviour achieved an F1-score of 81.31% and AUC of 81.27% [20]. Opinion mining using VADER with RF, SVC, and LSTM produced accuracies of 83.97%, 88.03%, and 94.71%, respectively [21]. Large language models and graph-based methods have also been explored for deceptive review detection. A dual-phase framework used LLM-generated synthetic reviews followed by ML/DL-based classification [22]. Aspect-based sentiment analysis combined with SenticNet and GCNs improved detection of human- and AI-generated fake reviews [23]. Psycholinguistic analysis with transformer models identified emotional exaggeration and cognitive complexity as useful deception indicators [24]. Positional encoding, BiLSTM, and hybrid attention were further shown to improve contextual review representation [25]. Multilingual fake review detection has been investigated using LLM-based data augmentation across multiple languages and domains, producing accuracy gains of up to 10.9% [26].

In a related low-resource setting, Roman Urdu fake review detection using textual, linguistic, and behavioural features showed that behavioural attributes significantly contributed to classification performance [27]. Behavioural and hybrid approaches have demonstrated that textual information alone may be insufficient. Joint modelling of review content, reviewer behaviour, and merchant behaviour improved fake review identification [28]. A hierarchical attention CNN with BERT tokenization achieved 91.2% accuracy and 93.2% F1-score [29], while conventional machine learning approaches on Yelp reviews also demonstrated reliable fake review classification [30]. Hybrid feature fusion has emerged as an effective strategy for combining complementary representations. BERT and BoW features integrated with RF, XGBoost, and stacking improved robustness by jointly considering contextual and behavioural information [31]. Structured metadata and semantic information were further combined for explainable fake review assessment [32], while linguistic and semantic feature analysis confirmed the importance of feature-level information for deceptive review classification [33].

RoBERTa-LSTM with attention and SHAP achieved 96.03% and 93.15% accuracy on two deceptive review datasets [34]. Similarly, integrating RoBERTa representations with reviewer behavioural features improved accuracy by approximately 3% [35].

Conventional machine learning classifiers also remained effective under different review data conditions [36]. Opinion-mining-based approaches have also contributed to fake review detection. TF-IDF with SVM and Naïve Bayes supported effective spam review verification [37], while comparative evaluation of six supervised classifiers demonstrated the applicability of conventional ML techniques [38]. An uncertainty-aware BERT framework achieved 91.75% accuracy [39], whereas sentiment and semantic information using SentiWordNet, WordNet, and deep learning achieved 92% accuracy [40]. Comparative studies have generally shown the advantage of contextual representation learning. Transformer models such as BERT and GPT-3 outperformed tree-based approaches including RF and XGBoost [41]. A BERT-SKEP-TextCNN framework achieved accuracies above 91% across multiple review domains [42]. Behaviour-sensitive feature extraction combined with contextual attention further demonstrated the complementary role of reviewer behaviour and textual content [43]. Finally, integrating multiple handcrafted feature categories has shown consistent benefits. Linguistic, POS, sentiment, and ontology-based features improved detection when used jointly [44]. A hybrid framework using 133 content-based and behavioural features with sampling-based class balancing achieved up to 89% accuracy, further supporting multidimensional feature fusion for fake review detection [45].

The reviewed studies demonstrate considerable progress through conventional machine learning [12], [17], [30], [36]–[38], deep learning and transformer representations [14]–[16], [19], [24], [25], [29], [34], [39], [41], [42], opinion mining [13], [21], [23], [40], and behavioural or hybrid feature modelling [20], [28], [31]–[35], [43]–[45]. However, the majority of existing approaches have been evaluated on English or other comparatively well-resourced review datasets. Although multilingual augmentation has been explored [26] and a related transliterated low-resource setting has been investigated for Roman Urdu [27], Roman Marathi–English code-mixed fake review detection remains insufficiently explored within the reviewed literature. Furthermore,1 existing studies demonstrate that semantic embeddings provide effective contextual representations, whereas linguistic, sentiment, behavioural, contextual, and metadata characteristics contain complementary indicators of deception. Nevertheless, the reviewed studies do not specifically investigate a unified framework combining LaBSE-based multilingual sentence semantics with opinion-mining and multidimensional handcrafted features for Roman Marathi code-mixed reviews. This motivates the proposed hybrid framework, in which LaBSE semantic embeddings are fused with linguistic, behavioural, contextual, temporal, demographic, metadata, and opinion-oriented features and subsequently evaluated using multiple machine learning classifiers for fake review detection in a low-resource Roman Marathi code-mixed environment.

3. MATERIALS AND METHODS

A. Roman Marathi Fake Review Dataset

Dataset Description

The proposed framework utilizes the Real-time Roman Marathi Fake Review Dataset, which was specifically developed to address the limited availability of annotated fake review datasets for Roman Marathi language content. Unlike publicly available benchmark datasets, this dataset was constructed through systematic collection and annotation of reviews obtained from multiple online platforms. The reviews are written primarily in Roman Marathi, where Marathi language is expressed using the Roman (Latin) alphabet together with commonly used English words and product-related expressions. This writing style closely resembles real user behaviour observed on contemporary e-commerce platforms and social media, making the dataset suitable for developing practical fake review detection models.

The complete dataset contains 2,287 annotated reviews, comprising 1,990 genuine reviews, 297 fake reviews, together with 34 descriptive attributes representing linguistic, behavioural, contextual, temporal, and metadata

information. The reviews cover a diverse range of product categories, review platforms, user profiles, and geographical locations, thereby enabling comprehensive evaluation of fake review detection under realistic multilingual review environments. Unlike conventional datasets that mainly focus on textual content, the proposed dataset integrates rich contextual and behavioural information that significantly enhances feature representation for machine learning models.

Data Acquisition

The dataset was developed through a systematic data acquisition process involving multiple online review sources. Reviews were collected from leading e-commerce platforms, including Amazon, Flipkart, Myntra, Ajio, and JioMart, together with social media platforms such as Facebook, Instagram, WhatsApp Groups, Telegram Groups, Discord Communities, and YouTube comments. Additional reviews were obtained from review websites, local business directories, and community discussion forums to capture diverse user opinions and writing behaviours. The collected reviews span the period 2023–2025, covering 45 cities across six Indian states, with Maharashtra serving as the primary linguistic region for Roman Marathi reviews. This diversified acquisition strategy ensures that the dataset represents various product domains, review platforms, and user communities while preserving the natural characteristics of real-world review data.

Each acquired review is associated with several complementary attributes, including reviewer information, platform characteristics, product details, linguistic properties, behavioural indicators, temporal information, geographical location, and authenticity-related metadata. These attributes provide valuable contextual information that supports comprehensive feature extraction and machine learning classification.

Data Validation and Annotation

Following data acquisition, the collected reviews undergo a rigorous validation and annotation process to ensure dataset quality and reliability. Initially, duplicate records, incomplete entries, formatting inconsistencies, and missing values are identified and corrected through automated quality assurance procedures. Subsequently, domain experts manually inspect each review and assign authenticity labels based on predefined annotation guidelines. Reviews are classified into Genuine and Fake categories by analysing linguistic consistency, reviewer behaviour, product-specific information, purchase verification, promotional content, grammatical quality, and contextual evidence.

After manual annotation, each review is enriched with a comprehensive set of descriptive features representing linguistic, behavioural, contextual, temporal, and metadata characteristics. These include sentiment-related information, emotional intensity, language confidence, grammar quality, authenticity indicators, platform information, reviewer demographics, temporal attributes, and product-specific information. The resulting dataset provides a reliable benchmark for developing and evaluating machine learning models for fake review detection in Roman Marathi review environments.

Dataset Attributes

Table 1 summarizes the principal variables collected during the data acquisition process and their respective purposes within the proposed framework.

Table 1: Acquired Variables of the Real-time Roman Marathi Fake Review Dataset

Feature Group	Representative Attributes
Textual	Review Text, Length, Word Count, Character Count
Linguistic	Language Confidence, Grammar Quality, Script
Opinion/Sentiment	Sentiment Score, Emotional Intensity, Rating
Behavioural	Helpfulness, User Type

Authenticity	Purchase Verified, Technical Details, Recommendation
Contextual	Platform, Product Category, Seller Response
Temporal/Demographic	Review Date/Time, Age Group, City/State

Dataset Characteristics

The developed dataset possesses several characteristics that distinguish it from existing fake review datasets:

 Reviews are collected from multiple online platforms, including e-commerce websites, social media, review platforms, and local business sources.

 Roman Marathi reviews preserve the natural writing behaviour commonly adopted by Indian users, including code-mixed Marathi and English expressions.

 The dataset contains 34 descriptive attributes that combine linguistic, behavioural, contextual, temporal, demographic, and authenticity-related information.

 Reviews span 70 product categories, 45 cities, 6 Indian states, and multiple reviewer profiles, providing substantial diversity for model development and evaluation.

 The dataset represents realistic fake review prevalence by preserving the natural imbalance between genuine and deceptive reviews while maintaining comprehensive feature annotations.

This comprehensive dataset serves as the foundation for implementing the proposed hybrid framework, enabling the integration of opinion mining, deep semantic feature extraction, handcrafted feature engineering, and machine learning-based classification for effective fake review detection in Roman Marathi review environments.

B. Proposed Hybrid Fake Review Detection Framework

The implementation of the proposed machine learning framework using the real-time Roman-Marathi fake review dataset is illustrated in Figure 1. Unlike the benchmark datasets presented in the previous sections, this implementation utilizes a self-developed Real-time Roman Marathi Fake Review Dataset consisting of Roman Marathi reviews collected from multiple online platforms. The implementation comprises data preprocessing, opinion mining, deep feature extraction, handcrafted feature engineering, feature fusion, machine learning-based classification, and performance evaluation.

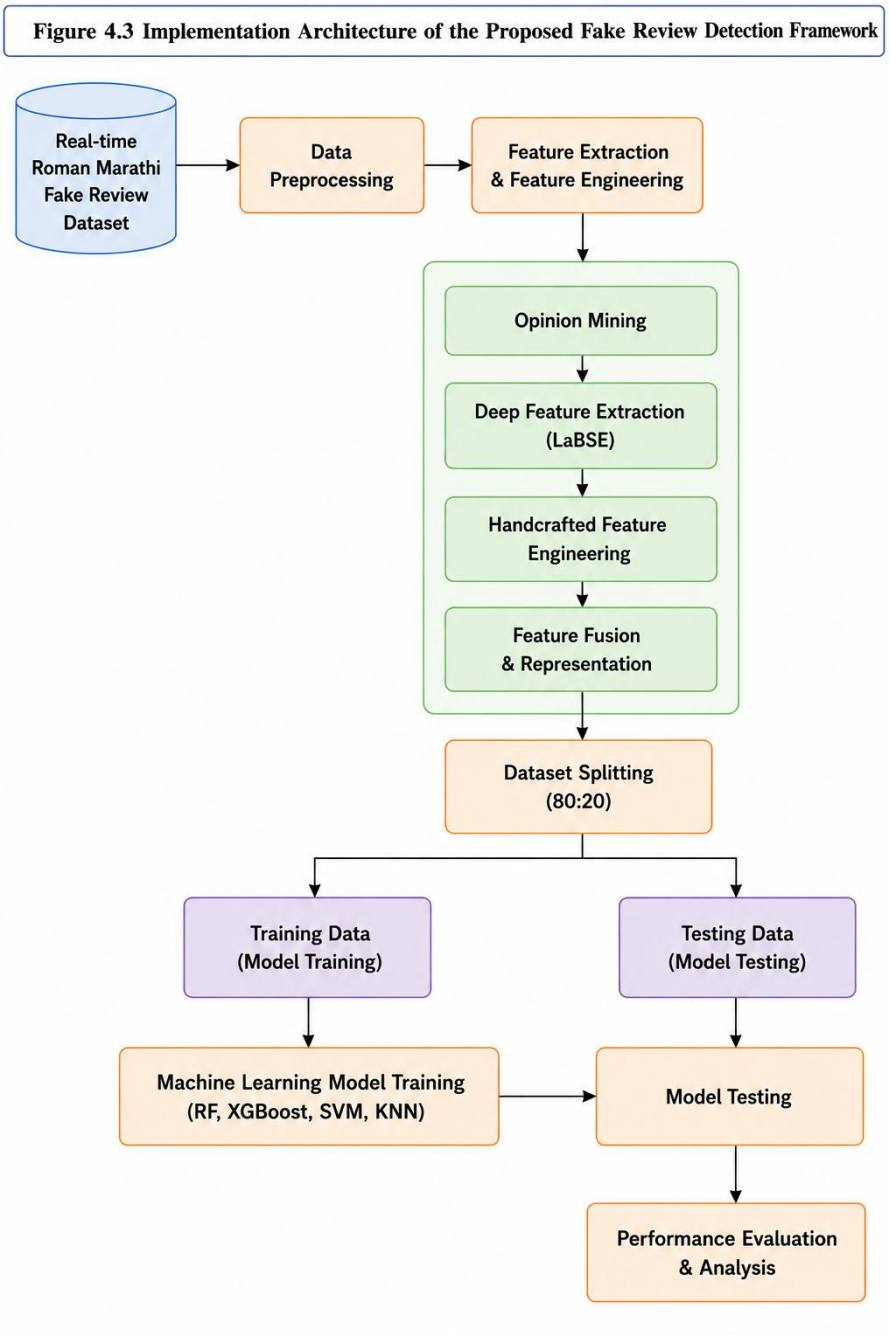

Figure 1: Proposed hybrid LaBSE and handcrafted feature fusion framework for Roman Marathi fake review detection.

Although the overall framework remains identical to the previous implementations, the deep feature extraction stage is specifically adapted for the linguistic characteristics of the collected review data. The review corpus contains Roman Marathi–English code-mixed text, where Marathi sentences are expressed using the Roman alphabet together with English vocabulary, product names, and platform-specific terminology. Consequently, the implementation employs a multilingual sentence embedding model to preserve semantic information more effectively than conventional English-only embedding models.

Data Preprocessing

The acquired review dataset undergoes a comprehensive preprocessing stage to transform raw review records into structured information suitable for feature extraction and machine learning classification. Initially, duplicate records, incomplete entries, missing values, and inconsistent attribute values are identified and removed to improve data quality. Numerical attributes are converted into appropriate data types, categorical variables are encoded into machine-readable representations, Boolean attributes are transformed into binary values, and temporal information is standardized to facilitate subsequent feature generation.

Unlike conventional English review datasets, the review text consists primarily of Roman Marathi expressions combined with English words, product names, and e-commerce terminology. Therefore, preprocessing is carefully designed to preserve the semantic meaning of Roman Marathi expressions while eliminating URLs, unnecessary symbols, excessive whitespace, and formatting inconsistencies. This strategy maintains the contextual integrity of code-mixed reviews and produces standardized textual representations suitable for opinion mining and multilingual semantic feature extraction.

Feature Extraction and Feature Engineering

Following preprocessing, informative features are generated through a combination of opinion mining, multilingual deep semantic feature extraction, handcrafted feature engineering, and feature fusion. The objective of this stage is to transform each review into a comprehensive numerical representation capable of distinguishing genuine reviews from deceptive reviews. Unlike relying solely on manually engineered features, the proposed implementation combines multilingual semantic representations with explicit linguistic, behavioural, contextual, and temporal information to improve classification performance.

4. Opinion Mining

Opinion mining is first performed to analyse the sentiment and contextual characteristics expressed within each review. Sentiment polarity, emotional intensity, helpfulness score, and authenticity-related indicators are extracted to characterize reviewer behaviour and linguistic patterns commonly associated with deceptive reviews.

Let the ith review be represented as Ti. The corresponding sentiment score is obtained as

$$S_i = \text{Sentiment}(T_i) \tag{1}$$

where Si denotes the sentiment polarity extracted from the review text.

Similarly, the emotional intensity associated with the review is represented as

$$E_i = \text{Emotion}(T_i) \tag{2}$$

where Ei represents the emotional characteristics of the review obtained during opinion mining.

5. LaBSE-Based Semantic Feature Extraction

The proposed implementation employs the Language-agnostic BERT Sentence Embedding (LaBSE) model to generate deep semantic representations from Roman Marathi review text. The present dataset consists of Roman Marathi–English code-mixed reviews, where Marathi expressions are written using the Roman alphabet together with English vocabulary and product-specific terminology. LaBSE was selected because its multilingual representation capability provides a suitable basis for modelling semantically heterogeneous and code-mixed review text.

Conventional English sentence embedding models are primarily optimized for monolingual English text and may not adequately preserve the semantic relationships present in multilingual code-mixed reviews. In contrast, LaBSE is specifically designed to learn language-independent sentence representations across multiple languages and mixed-language inputs. Consequently, LaBSE is more suitable for capturing the semantic characteristics of Roman Marathi reviews while simultaneously preserving the contextual relationships of embedded English expressions.

Let the preprocessed review text be represented as

$$T_i = \{w_1, w_2, \ldots, w_n\} \tag{3}$$

where wj denotes the jth token of the review.

The multilingual semantic embedding generated using LaBSE is expressed as

$$D_i = \text{LaBSE}(T_i) \tag{4}$$

where Di denotes the deep semantic representation of the ith review.

6. Handcrafted Feature Engineering

Besides semantic representations, several handcrafted features are extracted from the review text and associated metadata to describe linguistic, behavioural, contextual, temporal, and authenticity-related characteristics. These include fake score, fake probability, review rating, review length, word count, character count, language confidence, sentiment score, helpfulness score, purchase verification status, reviewer type, emotional intensity, grammar quality, technical details, recommendation indicator, seller response, device information, product category, and platform-related information.

The handcrafted feature vector corresponding to the ith review is represented as

$$H_i = [h_1, h_2, \ldots, h_m] \tag{5}$$

where hj represents the jth handcrafted feature extracted from the review and its associated metadata.

$$H_i = [F_{score}, F_{prob}, R, W, C, L, S, H, G, P, \ldots] \tag{6}$$

Here,

- *Fscore* = Fake Score
- *Fprob* = Fake Probability
- *R* = Rating
- *W* = Word Count
- *C* = Character Count
- *L* = Review Length
- *S* = Sentiment Score
- *H* = Helpfulness Score
- *G* = Grammar Quality
- *P* = Purchase Verification

These handcrafted features explicitly characterize reviewer behaviour and contextual information, thereby complementing the semantic representations learned by the deep learning model.

7. Hybrid Feature Fusion and Representation

The final representation of each review is obtained by integrating three complementary feature groups: the multilingual semantic embeddings generated by LaBSE, the opinion-mining features extracted from the review text, and the handcrafted linguistic, behavioural, contextual, temporal, and authenticity-related features.

Let

Di = [d1, d2,…, dp]

represent the LaBSE-based deep semantic feature vector,

Oi = [o1, o2,…, oq]

represent the opinion-mining feature vector containing attributes such as sentiment polarity, emotional intensity, helpfulness score, and authenticity-related opinion indicators, and

Hi = [h1, h2,…, hm]

represent the handcrafted feature vector containing linguistic, behavioural, contextual, temporal, demographic, platform-related, and metadata features.

The final hybrid feature representation is defined as

$$F_i = D_i \parallel O_i \parallel H_i \tag{7}$$

where $\parallel$ denotes vector concatenation.

The resulting feature vector combines implicit multilingual semantic information with explicit opinion-oriented and metadata-based characteristics. LaBSE captures contextual meaning in Roman Marathi–English code-mixed text, opinion-mining features represent the emotional and evaluative orientation of the review, and handcrafted features describe reviewer behaviour, review structure, platform context, temporal patterns, and authenticity indicators. This fusion provides a more comprehensive review representation than relying on any individual feature category and serves as the input to the selected machine learning classifiers.

Dataset Splitting

Following feature generation, the hybrid feature vectors are divided into training and testing subsets using an 80:20 stratified split to evaluate the generalization capability of the proposed framework. Stratified sampling preserves the original distribution of genuine and fake reviews within both subsets, thereby ensuring unbiased model evaluation.

Let the complete dataset be represented as D. The dataset partition is expressed as

$$D = D_{\text{train}} \cup D_{\text{test}} \tag{8}$$

subject to

$$|D_{\text{train}}| = 0.8|D|, \tag{9}$$
$$|D_{\text{test}}| = 0.2|D|.$$

The training subset is used to learn the classification models, whereas the testing subset remains completely unseen during training and is reserved for independent model evaluation.

Machine Learning Model Training

The generated hybrid feature vectors are supplied to four supervised machine learning algorithms, namely Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN). These classifiers are selected because they collectively represent ensemble learning, boosting, margin-based learning, and distance-based learning approaches, thereby enabling comprehensive comparative evaluation under identical experimental conditions.

Each classifier is independently trained using the generated hybrid feature vectors and subsequently employed to classify unseen reviews into Genuine and Fake categories.

8. Random Forest (RF)

Random Forest is an ensemble learning algorithm that constructs multiple decision trees using randomly selected subsets of the training data and feature space. During the training phase, each decision tree independently learns decision rules from different bootstrap samples of the hybrid feature vectors. The final prediction is obtained by aggregating the outputs of all decision trees through majority voting, thereby improving robustness and reducing overfitting. For a review represented by Fi, the predicted class is computed as

$$\hat{y}_i^{RF} = \text{mode}\left\{h_1(F_i), h_2(F_i), \ldots, h_M(F_i)\right\} \tag{10}$$

Where,

- $h_m(\cdot)$ denotes the prediction of the *mth* decision tree,
- M represents the total number of trees.

During testing, each unseen review is evaluated by all decision trees, and the majority class is assigned as the final prediction.

9. Support Vector Machine (SVM)

Support Vector Machine constructs an optimal separating hyperplane that maximizes the margin between genuine and fake review classes within the hybrid feature space. Since the extracted features are not always linearly separable, the Radial Basis Function (RBF) kernel is employed to project the data into a higher-dimensional space where nonlinear decision boundaries can be effectively learned. The SVM decision function is expressed as

$$\hat{y}_i^{SVM} = \text{sign}\left(w^T \phi(F_i) + b\right) \tag{11}$$

Where,

- $\phi(\cdot)$ represents the nonlinear kernel mapping,
- w denotes the learned weight vector,
- b is the bias term.

During testing, each review is assigned to the class determined by the sign of the decision function.

10. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting is an ensemble boosting algorithm that sequentially constructs decision trees to minimize the classification error of previous trees. During training, each newly generated tree learns to correct the residual errors produced by the preceding ensemble, thereby improving predictive performance. The prediction function is given by

$$\hat{y}_i^{XGB} = \sum_{m=1}^{M} \eta_m f_m(F_i) \tag{12}$$

Where,

- $f_m(\cdot)$ denotes the *mth* boosted decision tree,
- ηm represents the learning weight associated with the tree.

The final class label is obtained by applying a threshold to the aggregated prediction score.

11. K-Nearest Neighbour (KNN)

K-Nearest Neighbour is a non-parametric classification algorithm that assigns a class label according to the majority class among the k nearest neighbours in the hybrid feature space. During training, the feature vectors are stored without explicitly constructing a predictive model. During testing, the Euclidean distance between the query review and all training samples is computed, and the nearest neighbours determine the predicted class. The prediction is expressed as

$$\hat{y}_i^{KNN} = \text{mode}\left(N_k(F_i)\right) \tag{13}$$

Where,

- *Nk(Fi)* denotes the set of the *k* nearest neighbours of review *Fi.*

Model Testing

After completion of the training process, the developed machine learning models are evaluated using the independent testing dataset. Each unseen review undergoes the same preprocessing, opinion mining, deep feature extraction, and handcrafted feature engineering procedures employed during training to generate the corresponding hybrid feature representation. The generated feature vector is then supplied to the trained machine learning classifier to determine whether the review is genuine or fake. This testing procedure ensures that the classification models are evaluated under identical conditions while preserving the consistency of the proposed framework.

Let Fi denote the fused feature vector of the ith review. The predicted review label is obtained as

$$\hat{y}_i = f(F_i) \tag{14}$$

Where

- *y^i* represents the predicted class label of the ith review,
- *Fi* denotes the final hybrid feature vector obtained through feature fusion, and
- $f(\cdot)$ represents the selected machine learning classifier (Random Forest, XGBoost, Support Vector Machine, or K-Nearest Neighbour).

The predicted output is expressed as

$$\hat{y}_i = \begin{cases} 0, & \text{Genuine Review} \\ 1, & \text{Fake Review} \end{cases} \tag{15}$$

where 0 denotes a genuine review and 1 denotes a fake review.

Performance Evaluation

The final stage of the implementation involves evaluating the developed machine learning models using the independent testing dataset. Each classifier is evaluated independently, and the obtained results are subsequently compared to identify the most suitable machine learning model for detecting deceptive reviews in Roman Marathi code-mixed environments.

C. Proposed Algorithm

The implementation of the proposed fake review detection framework using the Real-time Roman Marathi Fake Review Dataset follows a systematic sequence of operations that integrates data acquisition, data validation, preprocessing, multilingual semantic feature extraction, handcrafted feature engineering, and supervised machine learning classification. Unlike the previous benchmark datasets, this implementation begins with the acquisition and annotation of real-time Roman Marathi reviews collected from multiple online platforms. The algorithm is designed to generate a comprehensive hybrid feature representation by combining opinion-based features, multilingual semantic

embeddings, and handcrafted metadata features before performing binary classification. The major steps of the proposed algorithm are summarized below.

Algorithm 1: Proposed Fake Review Detection Using the Real-time Roman Marathi Fake Review Dataset
Input: Real-time Roman Marathi Fake Review Dataset containing Roman Marathi review text, metadata, behavioural information, temporal attributes, and class labels. **Output:** Predicted review label (Genuine or Fake).
Step 1: Acquire review data from multiple online platforms and construct the Real-time Roman Marathi Fake Review Dataset. *Step 2:* Validate the collected reviews by removing duplicate, incomplete, and inconsistent records. *Step 3:* Annotate the collected reviews into Genuine and Fake categories according to predefined annotation guidelines. *Step 4:* Perform data preprocessing, including text cleaning, Roman Marathi text normalization, categorical encoding, Boolean conversion, temporal data standardization, and missing value handling. *Step 5:* Perform opinion mining to extract sentiment polarity, emotional intensity, helpfulness score, and authenticity-related opinion features. *Step 6:* Generate multilingual semantic embeddings from the Roman Marathi review text using the LaBSE model. *Step 7:* Extract handcrafted linguistic, behavioural, contextual, temporal, and metadata features from the review records. *Step 8:* Fuse the opinion-mining features, LaBSE semantic embeddings, and handcrafted features to construct the final hybrid feature vector. *Step 9:* Split the dataset into training and testing subsets using an 80:20 stratified sampling strategy. *Step 10:* Train the selected machine learning classifiers (Random Forest, XGBoost, Support Vector Machine, and K-Nearest Neighbour) using the training dataset. *Step 11:* Classify the testing reviews into Genuine and Fake categories using the trained machine learning models. *Step 12:* Store the predicted labels for subsequent performance evaluation and comparative analysis.

12. EXPERIMENTAL RESULT AND DISCUSSION

A. Experimental Setup

The experimental evaluation of the proposed framework was conducted using a Python-based machine learning environment to ensure consistent implementation across all three review datasets. The experiments included data preprocessing, opinion mining, deep feature extraction, handcrafted feature engineering, feature fusion, machine learning model training, and performance evaluation.

The implementation was carried out on a computer system running the Windows 11 operating system equipped with an Intel Core i5 processor, 16 GB RAM, and an NVIDIA GeForce RTX 3050 GPU. The software environment consisted of Python 3.8, developed using the PyCharm Integrated Development Environment (IDE) with the Anaconda distribution for package and environment management. This configuration provided sufficient computational resources for generating transformer-based semantic embeddings, training multiple machine learning classifiers, and evaluating the proposed framework efficiently.

Each dataset underwent its corresponding preprocessing and feature engineering pipeline before being partitioned into 80% training and 20% testing subsets using stratified sampling wherever applicable. The training data were used for model learning and hyperparameter optimization, while the testing data were reserved exclusively for evaluating the generalization capability of the trained classifiers.

The Language-agnostic BERT Sentence Embedding (LaBSE) model was utilized for the Real-time Roman Marathi Fake Review Dataset owing to its superior capability in representing multilingual and code-mixed textual content. The extracted semantic embeddings were fused with opinion-mining and handcrafted features before being supplied to the machine learning classifiers. Four supervised machine learning algorithms, namely Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN), were independently trained and evaluated using identical experimental settings. This comparative evaluation enabled the identification of the most effective classifier for fake review detection under different review environments.

The hardware and software configuration used throughout the experimental study is summarized in Table 2.

Table 2: Hardware and Software Configuration

Parameter	**Specification**
Operating System	Windows 11 (64-bit)
Processor	Intel Core i5
RAM	16 GB
Graphics Processing Unit	NVIDIA GeForce RTX 3050
Programming Language	Python 3.8
IDE	PyCharm
Python Distribution	Anaconda
Deep Feature Extraction	Sentence-BERT (English datasets), LaBSE (Roman Marathi dataset)
Train-Test Split	80 : 20
Machine Learning Models	RF, XGBoost, SVM, KNN

B. Performance Evaluation Metrics

The effectiveness of the proposed fake review detection framework was evaluated using widely accepted performance metrics for binary classification. Since the objective of the proposed system is to classify reviews into genuine and fake categories, multiple evaluation measures were employed to comprehensively assess the classification performance from different perspectives.

The evaluation was performed independently for each of the three review datasets using the same set of performance measures. These metrics quantify the predictive capability of the selected machine learning classifiers while facilitating a fair comparison across different review environments.

The confusion matrix serves as the foundation for computing all evaluation metrics. It summarizes the relationship between actual review labels and predicted review labels by categorizing predictions into True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

- *True Positive (TP):* Fake reviews correctly classified as fake.
- *True Negative (TN):* Genuine reviews correctly classified as genuine.
- *False Positive (FP):* Genuine reviews incorrectly classified as fake.

- *False Negative (FN):* Fake reviews incorrectly classified as genuine.

Based on the confusion matrix, the following evaluation metrics were computed.

Accuracy

Accuracy measures the overall proportion of correctly classified reviews among all reviews in the testing dataset. It provides a general indication of the classifier's prediction capability.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

Precision

Precision evaluates the proportion of reviews predicted as fake that are actually fake. It reflects the reliability of positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

Recall

Recall, also referred to as Sensitivity or True Positive Rate, measures the capability of the classifier to correctly identify fake reviews.

$$Recall = \frac{TP}{TP + FN} \quad (18)$$

F1-Score

The F1-score represents the harmonic mean of precision and recall. It provides a balanced evaluation when both false positives and false negatives are important.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (19)$$

Confusion Matrix

The confusion matrix provides a detailed visualization of classification outcomes by comparing predicted labels with actual labels. It enables analysis of correctly classified reviews as well as the distribution of classification errors.

In this research, the confusion matrix was generated separately for each machine learning classifier and for each review dataset. The matrix serves as the primary tool for analysing model behaviour and interpreting the experimental results obtained from the proposed fake review detection framework.

C. Experimental Results Using Real-time Roman Marathi Fake Review Dataset

To evaluate the practical applicability of the proposed hybrid fake review detection framework, experiments were conducted using a real-time Roman Marathi fake review dataset containing reviews collected from multiple online platforms. After performing data validation, duplicate removal, and preprocessing, the dataset was analyzed to understand its class distribution, platform diversity, and textual characteristics before feature extraction and classification. Exploratory Data Analysis (EDA) was carried out to visualize the overall dataset composition and identify patterns between genuine and fake reviews. The resulting observations provide valuable insights into the nature of multilingual Roman Marathi reviews and justify the need for a robust hybrid feature representation combining opinion mining, multilingual deep learning embeddings, and handcrafted linguistic features.

Dataset Analysis

This section presents the exploratory analysis of the Real-time Roman Marathi Fake Review Dataset before model training. The dataset contains both genuine and fake reviews collected from multiple online sources, providing a realistic representation of Roman Marathi user-generated content. The analysis focuses on the class distribution, rating distribution, platform-wise review distribution, and textual characteristics of both review categories.

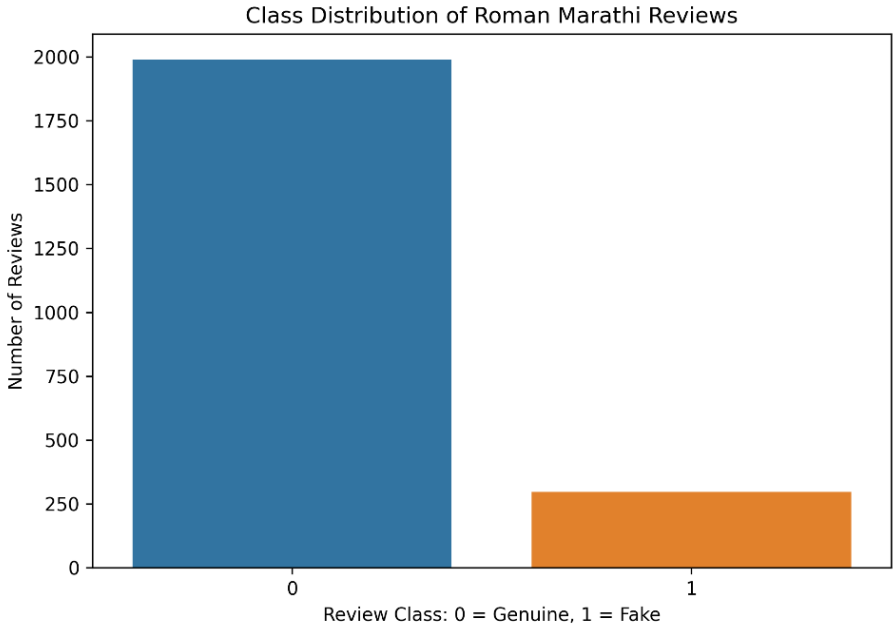

Figure 2: Class Distribution of the Real-time Roman Marathi Fake Review Dataset

Figure 2: illustrates the distribution of genuine and fake reviews in the collected Real-time Roman Marathi Fake Review Dataset. A total of 2,287 reviews are included, consisting of 1,990 genuine reviews (87.01%) and 297 fake reviews (12.99%). The figure shows that genuine reviews constitute the majority of the dataset, while fake reviews represent a comparatively smaller portion. This uneven distribution indicates the presence of class imbalance, which may bias the learning process of machine learning models. Therefore, a dataset balancing technique is applied before model training to ensure equal representation of both classes and improve the reliability of the classification results.

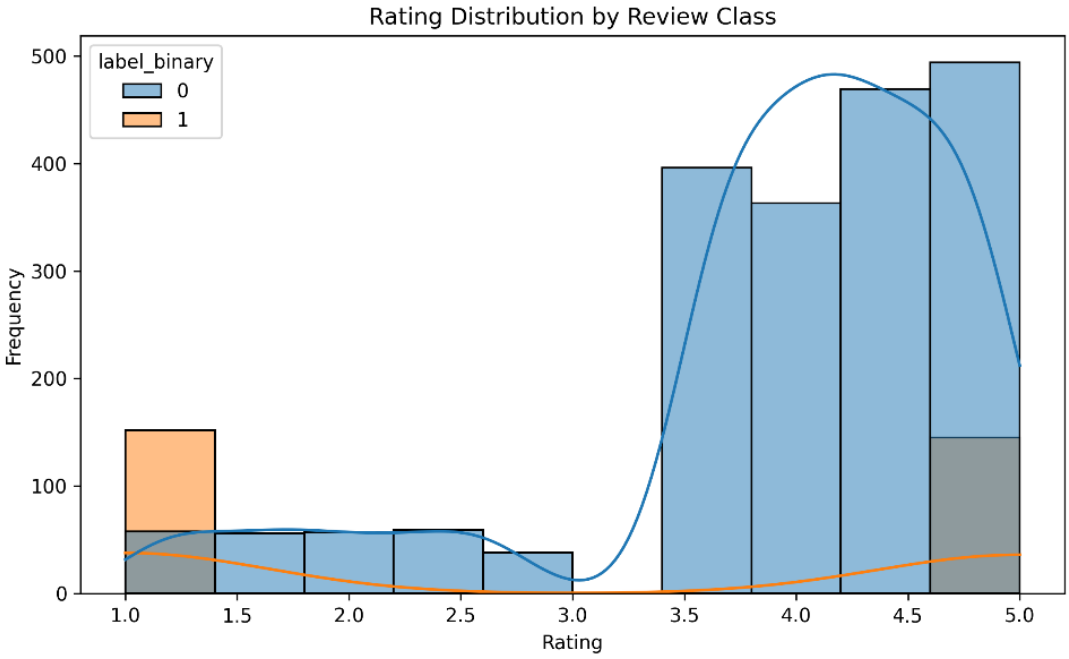

Figure 3: Review Distribution Across Different Online Platforms

Figure 3: presents the distribution of review ratings for genuine and fake reviews. Genuine reviews are primarily concentrated between ratings 4 and 5, indicating that authentic users generally provide positive feedback based on actual experiences. In contrast, fake reviews exhibit a wider variation, with many reviews assigned the lowest rating as well as a noticeable concentration of five-star ratings. This behaviour suggests that fraudulent reviews are often written to either excessively promote or deliberately criticize products and services.

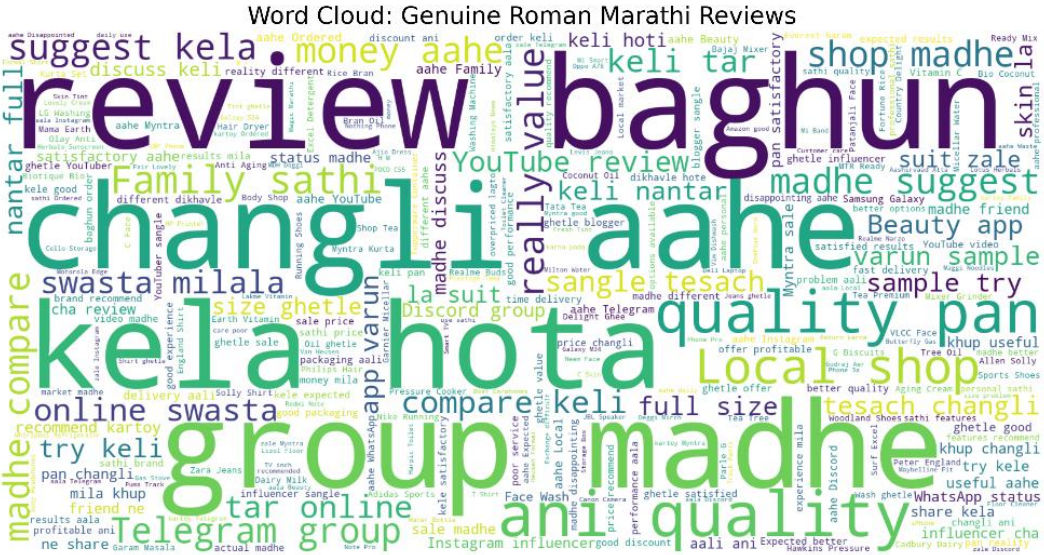

Figure 4: Word Cloud of Genuine Roman Marathi Reviews

To further investigate the textual characteristics of the collected reviews, separate word clouds were generated for genuine and fake reviews. The genuine review word cloud, shown in Figure 4, contains a wide variety of product-related, experience-based, and descriptive terms such as quality, baghun, review, family, money, shop, and recommend. These words indicate that genuine reviews generally contain richer contextual information, detailed

product descriptions, purchasing experiences, and balanced opinions. The higher lexical diversity observed in genuine reviews reflects natural human writing behaviour and provides meaningful semantic information for feature extraction.

Figure 5: Word Cloud of Fake Roman Marathi Reviews

In contrast, the fake review word cloud shown in Figure 5 reveals a noticeably different linguistic pattern. Frequently occurring words such as Amazing, Perfect, Excellent, Must buy, Recommended, and Working dominate the visualization, highlighting the excessive use of highly positive promotional language. At the same time, several strongly negative expressions, including Worst, Bad quality, Money waste, and Terrible, are also repeatedly observed. Compared with genuine reviews, fake reviews tend to employ repetitive sentiment-oriented vocabulary, exaggerated emotional expressions, and less descriptive product-specific information. These textual differences demonstrate the importance of combining multilingual semantic embeddings with handcrafted linguistic and opinion-mining features for effective fake review detection in Roman Marathi text.

Overall, the exploratory analysis confirms that the proposed dataset reflects realistic online review characteristics, including class imbalance, heterogeneous data sources, and significant linguistic differences between genuine and fake reviews. These observations establish a strong foundation for the subsequent feature extraction, feature fusion, and machine learning-based classification stages of the proposed framework.

Dataset Balancing

This section discusses the balancing of the Real-time Roman Marathi Fake Review Dataset prior to model training. The original dataset contains 1,990 genuine reviews and 297 fake reviews, resulting in an imbalanced class distribution. To reduce classification bias and improve the learning capability of the machine learning models, the Random Oversampling technique was applied to the minority class by increasing the number of fake reviews to match the majority class. Consequently, the balanced dataset consists of 1,990 genuine reviews and 1,990 fake reviews, producing a total of 3,980 reviews. The balanced dataset was subsequently divided using an 80:20 train-test split, resulting in 3,184 training samples and 796 testing samples. This balanced training strategy ensures equal representation of both review classes, enabling the classifiers to learn discriminative patterns more effectively and providing a more reliable and unbiased evaluation of the proposed fake review detection framework.

Confusion Matrix Analysis

This section presents the confusion matrix analysis of the Random Forest, Support Vector Machine, XGBoost, and K-Nearest Neighbour classifiers evaluated using the balanced Real-time Roman Marathi Fake Review Dataset. The balanced test set contains 398 genuine reviews and 398 fake reviews, resulting in a total of 796 test samples. The

confusion matrices indicate the number of correctly and incorrectly classified instances for both review classes and help evaluate the error patterns of each classifier.

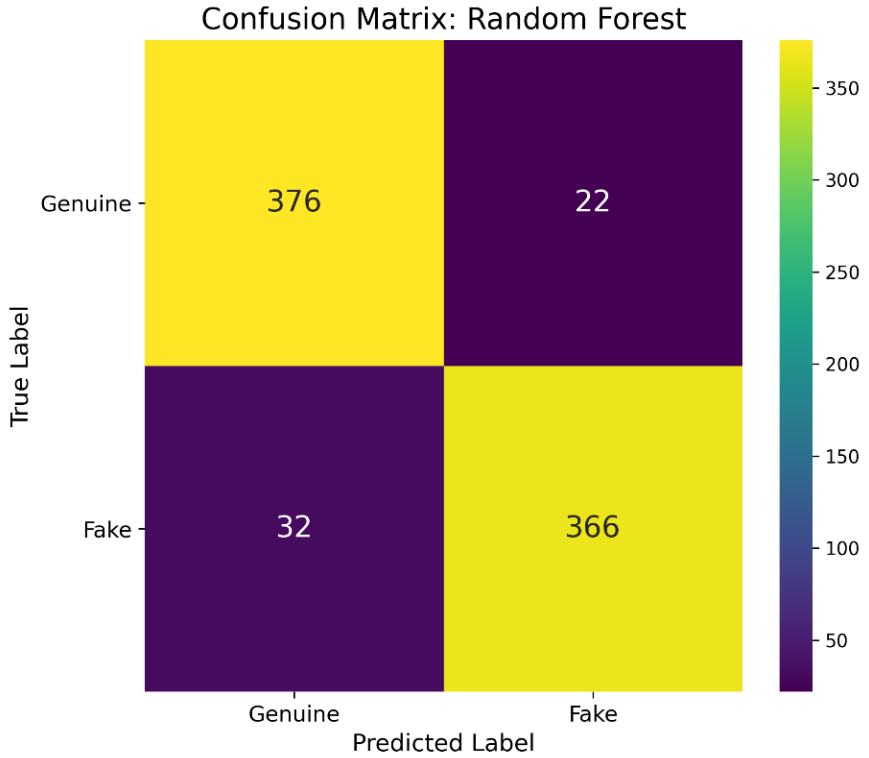

Figure 6: Confusion Matrix of the Random Forest Classifier for Roman Marathi Fake Review Detection

Figure 6: presents the confusion matrix obtained using the Random Forest classifier. Out of 398 genuine reviews, 376 were correctly classified, whereas 22 were incorrectly identified as fake. Similarly, the classifier correctly detected 366 fake reviews, while 32 fake reviews were misclassified as genuine. Thus, Random Forest correctly classified 742 out of 796 reviews, achieving an overall accuracy of approximately 93.22%. The smaller number of false positives than false negatives indicates that the classifier was comparatively more precise in predicting fake reviews but missed some fraudulent instances.

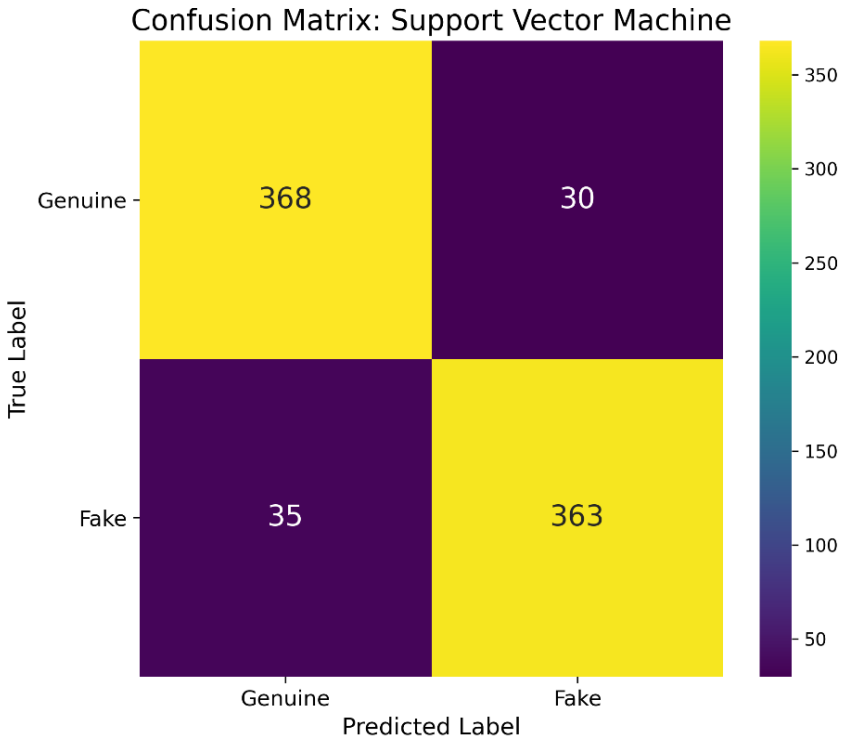

Figure 7: Confusion Matrix of the Support Vector Machine Classifier for Roman Marathi Fake Review Detection

Figure 7: illustrates the classification results of the Support Vector Machine. The model correctly recognized 368 genuine reviews and 363 fake reviews, while 30 genuine reviews were incorrectly classified as fake and 35 fake reviews were incorrectly classified as genuine. Overall, SVM correctly classified 731 out of 796 reviews, resulting in an accuracy of approximately 91.83%. The similar numbers of false positives and false negatives demonstrate relatively balanced classification behaviour across both classes.

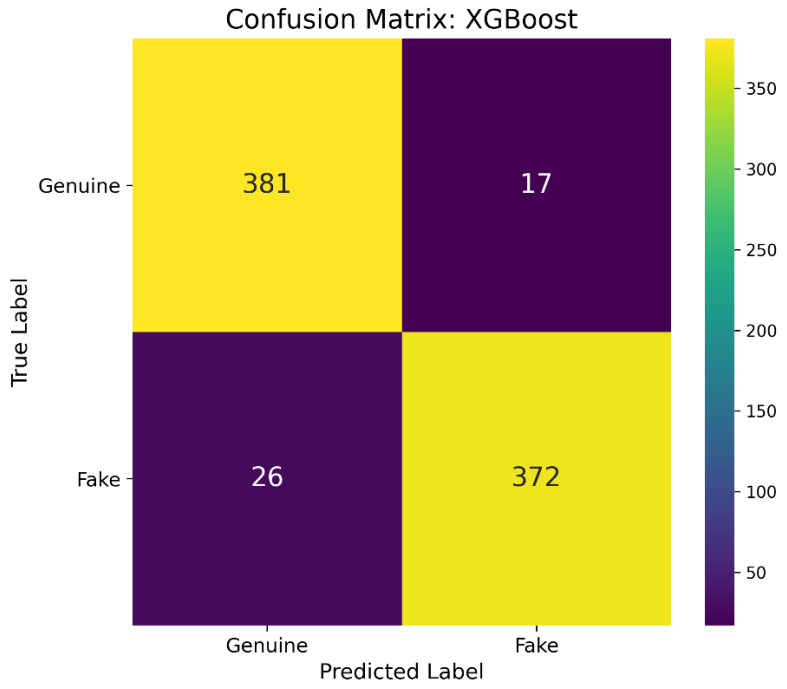

Figure 8: Confusion Matrix of the XGBoost Classifier for Roman Marathi Fake Review Detection

Figure 8: shows the confusion matrix of the XGBoost classifier. The model correctly classified 381 genuine reviews and 372 fake reviews. Only 17 genuine reviews were incorrectly predicted as fake, while 26 fake reviews were misclassified as genuine. XGBoost correctly identified 753 out of 796 test reviews, producing the highest accuracy of approximately 94.60% among the evaluated classifiers. Its lower number of overall errors demonstrates its stronger capability to learn complex relationships from the fused multilingual, opinion-based, and handcrafted features.

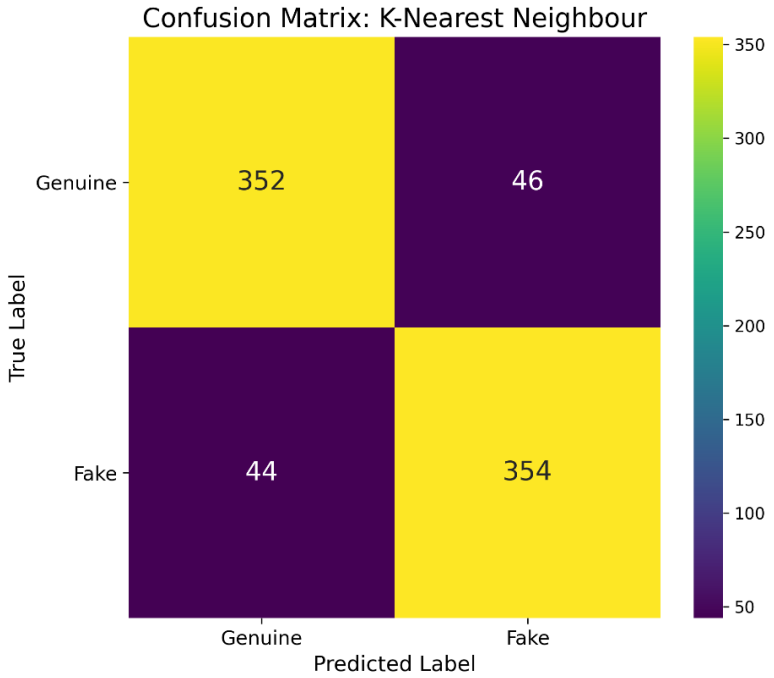

Figure 9: Confusion Matrix of the K-Nearest Neighbour Classifier for Roman Marathi Fake Review Detection

Figure 9: presents the confusion matrix of the K-Nearest Neighbour classifier. Among the genuine reviews, 352 were correctly classified and 46 were incorrectly identified as fake. For the fake-review class, 354 reviews were correctly detected, whereas 44 were incorrectly predicted as genuine. The classifier correctly categorized 706 out of 796 reviews, resulting in an overall accuracy of approximately 88.69%. The comparatively larger number of classification errors indicates that KNN is more sensitive to local feature similarity and may be less effective when applied to the high-dimensional fused feature representation.

Overall, the confusion matrix results indicate that XGBoost achieved the best classification performance, followed by Random Forest, Support Vector Machine, and K-Nearest Neighbour. XGBoost produced the lowest number of misclassified reviews, whereas KNN generated the highest number of errors. These findings demonstrate that ensemble-based classifiers are more effective for identifying complex linguistic and behavioural patterns in Roman Marathi fake reviews.

Overall Comparative Results Analysis

This section presents the overall comparative performance of the four machine learning classifiers evaluated using the balanced Real-time Roman Marathi Fake Review Dataset. The comparison is based on Accuracy, Precision, Recall, and F1-score, which collectively indicate the models' overall classification capability, reliability in identifying fake reviews, sensitivity toward fraudulent instances, and balance between precision and recall. The comparative results are presented in Table 3.

Table 3: Overall Performance Comparison for Roman Marathi Fake Review Detection

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	93.22	94.33	91.96	93.13
Support Vector Machine	91.83	92.37	91.21	91.78
XGBoost	94.60	95.63	93.47	94.54
K-Nearest Neighbour	88.69	88.50	88.94	88.72

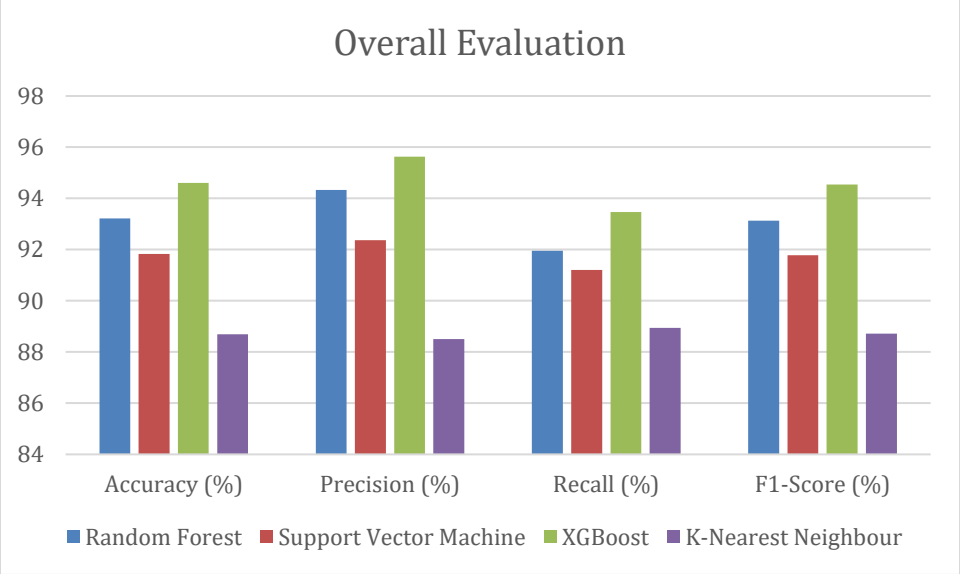

Figure 10: Overall Performance Evaluation for Roman Marathi Fake Review Detection

As shown in Figure 10, XGBoost achieved the best overall performance, with an accuracy of 94.60%, precision of 95.63%, recall of 93.47%, and F1-score of 94.54%. Its higher precision indicates that most reviews classified as fake were correctly identified, while its recall confirms that the model successfully detected a large proportion of the actual fake reviews. The lower number of false positives and false negatives observed in its confusion matrix further supports its superior classification performance.

Random Forest obtained the second-best results, achieving 93.22% accuracy and an F1-score of 93.13%. Its precision of 94.33% was higher than its recall of 91.96%, indicating that the classifier was reliable when predicting a review as fake but failed to detect some fraudulent reviews. The Support Vector Machine achieved comparatively balanced results, with 91.83% accuracy, 92.37% precision, 91.21% recall, and a 91.78% F1-score.

K-Nearest Neighbour produced the lowest performance among the evaluated classifiers, with an accuracy of 88.69% and an F1-score of 88.72%. This comparatively lower result can be associated with the sensitivity of distance-based classification to the high-dimensional fused feature space generated from LaBSE embeddings, opinion-mining features, and handcrafted attributes.

Overall, the results confirm that the proposed hybrid representation effectively supports the classification of Roman Marathi fake reviews despite variations in spelling, code-mixed language, informal sentence construction, and platform-specific writing styles. The strong performance of XGBoost demonstrates its ability to learn complex non-linear relationships among multilingual semantic, opinion-based, linguistic, behavioural, contextual, and temporal features. Therefore, XGBoost is identified as the most suitable classifier for the proposed Real-time Roman Marathi Fake Review Detection framework.

To further contextualize the effectiveness of the proposed framework, its performance was compared with selected existing fake review detection approaches that employ semantic representations, sentiment information, behavioural features, or hybrid feature learning. Since these studies use different datasets, languages, data distributions, and evaluation protocols, the comparison is intended to indicate the relative performance and methodological characteristics rather than establish a direct experimental superiority.

Table 4: Comparative Analysis of the Proposed Framework with Existing Fake Review Detection Approaches

Ref.	Method / Feature Representation	Dataset / Language	Best Model / Approach	Accuracy (%)	F1-Score (%)
[27]	Textual + linguistic + behavioural features with data balancing	Roman Urdu reviews	Gradient Boosting	~91.00	–
[29]	BERT tokenization + hierarchical attention CNN	Fake review datasets / English	HACNN	91.20	93.20
[45]	133 content-based + behavioural features with sampling	Yelp / English	Hybrid feature-based CNN	89.00	–
Proposed	LaBSE semantic + opinion-mining + handcrafted feature fusion	Real-time Roman Marathi–English code-mixed reviews	XGBoost	94.60	94.54

Table 4 presents a comparative analysis of the proposed framework with selected existing fake review detection approaches that are closely related in terms of low-resource language processing, semantic representation, behavioural features, and hybrid feature learning. Since the studies employ different datasets, languages, feature representations, and experimental settings, the comparison is intended to provide a contextual performance assessment rather than a direct model-to-model evaluation.

Among the existing approaches, the Roman Urdu-based study combined textual, linguistic, and behavioural features with data balancing and achieved approximately 91.00% accuracy using Gradient Boosting [27]. This study is particularly relevant because it addresses fake review detection in a transliterated regional-language setting. The HACNN approach employed BERT tokenization with hierarchical attention-based feature learning and achieved 91.20% accuracy and 93.20% F1-score on English fake review datasets [29]. In comparison, the hybrid feature-based CNN incorporating 133 content-based and behavioural features with sampling achieved an accuracy of 89.00% on the English Yelp dataset [45].

The proposed framework achieved 94.60% accuracy and 94.54% F1-score using XGBoost, representing the highest reported accuracy among the approaches considered in Table 4. Compared with the approximately 91.00% accuracy reported for Roman Urdu using Gradient Boosting [27], the obtained results indicate the potential benefit of combining multilingual semantic information with explicit review characteristics for transliterated low-resource language processing. The proposed framework also achieved higher accuracy than HACNN [29] and the hybrid content- and behaviour-based approach [45]. Moreover, its F1-score of 94.54% is higher than the 93.20% reported for HACNN [29], indicating a favourable balance between fake-review precision and recall.

The comparative results suggest that the integration of LaBSE-based semantic embeddings, opinion-mining features, and handcrafted linguistic, behavioural, contextual, temporal, and metadata information provides a comprehensive representation of Roman Marathi–English code-mixed reviews. Unlike the selected approaches that primarily emphasize behavioural and linguistic features [27], hierarchical contextual learning [29], or content-behaviour feature integration [45], the proposed framework combines multilingual sentence-level semantics with multiple explicit feature categories within a unified feature-fusion strategy.

Nevertheless, the reported numerical improvements should be interpreted carefully because the compared studies were evaluated using different datasets and linguistic environments. The proposed study specifically addresses real-time Roman Marathi–English code-mixed reviews, which contain transliteration, spelling variations, informal expressions, and mixed-language usage. Therefore, rather than establishing direct superiority over existing studies, the results demonstrate the effectiveness of the proposed hybrid feature representation and establish a performance benchmark for Roman Marathi code-mixed fake review detection.

13. CONCLUSION

This paper presented a hybrid LaBSE semantic and handcrafted feature fusion framework with machine learning for fake review detection in Roman Marathi code-mixed text. The proposed approach integrated opinion-mining features, LaBSE-based multilingual semantic embeddings, and handcrafted linguistic, behavioural, contextual, temporal, demographic, and metadata features to provide a comprehensive representation of genuine and deceptive reviews. The developed dataset consisted of 2,287 Roman Marathi reviews collected from multiple online platforms, and random oversampling was applied to address class imbalance before model training and evaluation. The fused feature representation was evaluated using Random Forest, XGBoost, SVM, and KNN classifiers. Experimental results showed that XGBoost achieved the best overall performance with 94.60% accuracy, 95.63% precision, 93.47% recall, and 94.54% F1-score, followed by Random Forest with 93.22% accuracy. These results demonstrate that combining multilingual semantic representations with explicit handcrafted and opinion-oriented features can effectively capture deceptive patterns in low-resource Roman Marathi code-mixed reviews.

The study provides an initial foundation for automated fake review detection in Roman Marathi and demonstrates the applicability of hybrid feature fusion for low-resource and transliterated review environments. Future work can focus on expanding the dataset across additional platforms and domains, investigating advanced multilingual transformer models, and evaluating cross-domain and cross-lingual generalization. Further research can also consider explainable AI and real-time deployment to improve the interpretability and practical applicability of fake review detection systems.

REFERENCES

1. A. Singh and T. Narang, “Deep learning approaches for fake review detection: A comprehensive survey,” Journal of Interdisciplinary Knowledge, vol. 8, p. e01631, Nov. 2025, doi: 10.37497/jik.v8iknowledge.1631.
2. R.-A. Duma, Z. Niu, A. S. Nyamawe, and A. A. Manjotho, “An analysis of graph neural networks for fake review detection: A systematic literature review,” Neurocomputing, vol. 623, Art. no. 129341, 2025, doi: 10.1016/j.neucom.2025.129341.
3. R. Gupta, I. Kashyap, and V. Jindal, “A comprehensive survey on fake review detection system with future directions,” in Lecture Notes in Networks and Systems, 2024, pp. 1–14, doi: 10.1007/978-981-97-4860-0_1.
4. R. Gupta, V. Jindal, and I. Kashyap, “Recent state-of-the-art of fake review detection: A comprehensive review,” Knowledge Engineering Review, vol. 39, p. e8, 2024, doi: 10.1017/S0269888924000067.
5. R.-A. Duma, A. Oproescu, T. Rebedea, and D. Mladenić, “Fake review detection techniques, issues, and future research directions: A literature review,” Knowledge and Information Systems, vol. 66, pp. 1–42, 2024, doi: 10.1007/s10115-024-02118-2.
6. K. Mane and S. Dongre, “A review of different machine learning techniques for fake review identification,” in Proc. 5th Int. Conf. Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 2024, pp. 476–481, doi: 10.1109/ICDICI62993.2024.10810890.
7. M. Jalal and M. Hussein, “Fake reviews detection in e-commerce using machine learning techniques: A comparative survey,” BIO Web of Conferences, vol. 97, Art. no. 00099, 2024, doi: 10.1051/bioconf/20249700099.

8. S. J. Bagastio, S. Suakanto, and H. Fakhrurroja, “A systematic analysis of machine learning and deep learning strategies for identifying fake reviews,” in Proc. Int. Conf. Intelligent Cybernetics Technology and Applications (ICICyTA), Bali, Indonesia, 2024, pp. 239–244, doi: 10.1109/ICICYTA64807.2024.10913161.
9. M. Ennaouri and A. Zellou, “Machine learning approaches for fake reviews detection: A systematic literature review,” Journal of Web Engineering, vol. 22, no. 5, pp. 821–848, Jul. 2023, doi: 10.13052/jwe1540-9589.2254.
10. S. K. Maurya, D. Singh, and A. K. Maurya, “Deceptive opinion spam detection approaches: A literature survey,” Applied Intelligence, vol. 53, no. 2, pp. 2189–2234, 2023, doi: 10.1007/s10489-022-03427-1.
11. H. AbouGrad and A. Shabarshov, “AI-powered fraud e-commerce reviews detection framework using machine learning hybrid neural network model,” in Artificial Intelligence Applications for Human Well-Being and a Sustainable Future, M. Sanmugam, Ed. Cham, Switzerland: Springer, 2026, ch. 15, doi: 10.1007/978-3-032-14373-0.
12. C. H. Vasanth Kumar and P. Krishnamurthy, “Fraud detection in online consumer reviews using stack ensemble,” AIP Conference Proceedings, vol. 3369, no. 1, Art. no. 060007, Mar. 2026, doi: 10.1063/5.0319331.
13. K. T., V. C., U. Surya, and V. S. B., “Psychological-based opinion mining for fake review detection using advanced NLP and machine learning techniques,” in Proc. Global Conf. Emerging Technology (GINOTECH), Pune, India, 2025, pp. 1–6, doi: 10.1109/GINOTECH63460.2025.11077032.
14. M. J. Abd and M. H. Hussein, “Detecting fake reviews in e-commerce using deep learning techniques,” in Proc. Int. Conf. Intelligent Systems and Computational Networks (ICISCN), Bidar, India, 2025, pp. 1–4, doi: 10.1109/ICISCN64258.2025.10934517.
15. J. M. T. Jayasinghe and S. Dassanayaka, “Detecting deception: Employing deep neural networks for fraudulent review detection on Amazon,” Neural Computing and Applications, vol. 37, pp. 21715–21742, 2025, doi: 10.1007/s00521-025-11485-y.
16. M. Kumar, A. Rana, A. K. Yadav, and D. Yadav, “Leveraging sentiment analysis to detect fake reviews using deep learning,” SN Computer Science, vol. 6, no. 3, 2025, doi: 10.1007/s42979-025-03792-x.
17. M. M. B. Alsaad and H. Joshi, “A framework for online fake review detection on Yelp electronics product dataset using machine learning techniques,” in Proc. International Conference on Next-Generation Communication and Computing (NGCCOM 2024), Lecture Notes in Networks and Systems, vol. 1305, Singapore: Springer, 2025, doi: 10.1007/978-981-96-3725-6_33.
18. R. Sharma, V. Sharma, T. K. Vashishth, S. Shashi, A. Pandey, and S. Chaudhary, “Revealing the reliability of Amazon products via innovative fake review detection using machine learning,” in Proc. 6th Int. Conf. Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2025, pp. 217–221, doi: 10.1109/ICICV64824.2025.11086089.
19. S. Geetha, E. Elakiya, R. S. Kanmani, and M. K. Das, “High performance fake review detection using pretrained DeBERTa optimized with Monarch Butterfly paradigm,” Scientific Reports, vol. 15, no. 1, Art. no. 7445, 2025, doi: 10.1038/s41598-025-89453-8.
20. J. Wang, J. Chen, and W. Zhang, “WF-CFRB: A deep learning approach for fake review detection based on weighted fusion of contextual features and reviewer behaviors,” Journal of Systems Science and Systems Engineering, vol. 34, pp. 558–575, 2025, doi: 10.1007/s11518-025-5641-4.
21. M. Chinta, J. Machcha, C. Chagamreddy, and M. Shaik, “Opinion-mining based fake review detection for genuine online products,” in Proc. Int. Conf. Computing Technologies and Data Communication (ICCTDC), Hassan, India, 2025, pp. 1–6, doi: 10.1109/ICCTDC64446.2025.11158925.
22. D. Nawara and R. Kashef, “A dual-phase framework for detecting authentic and computer-generated customer reviews using large language models,” Decision Analytics Journal, vol. 15, Art. no. 100581, 2025, doi: 10.1016/j.dajour.2025.100581.
23. P. Phukon et al., “Detecting fake reviews using aspect-based sentiment analysis and graph convolutional networks,” Applied Sciences, vol. 15, no. 7, Art. no. 3771, 2025, doi: 10.3390/app15073771.
24. J. Salminen, M. Mustak, S.-G. Jung, et al., “Decoding deception in the online marketplace: Enhancing fake review detection with psycholinguistics and transformer models,” Journal of Marketing Analytics, 2025, doi: 10.1057/s41270-025-00393-8.
25. J. Chen, T. Zhang, Z. Yan, et al., “Attention-based BiLSTM with positional embeddings for fake review detection,” Journal of Big Data, vol. 12, Art. no. 83, 2025, doi: 10.1186/s40537-025-01130-9.
26. M. Liu and M. Poesio, “Data augmentation for fake reviews detection in multiple languages and multiple domains,” arXiv preprint arXiv:2504.06917, 2025.
27. N. Mughal, G. Mujtaba, M. H. Mughal, A. Manaf, and Z. Kamangar, “Fake reviews detection on e-commerce websites using novel user behavioral features: An experimental study,” ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 24, no. 9, pp. 1–44, Sep. 2025, doi: 10.1145/3748493.
28. P. Sun et al., “Fake review detection model based on comment content and review behavior,” Electronics, vol. 13, no. 21, Art. no. 4322, 2023, doi: 10.3390/electronics13214322.
29. B. P. Sahoo and S. K. Rath, “HACNN: Hierarchical attention convolutional neural network for fake review detection,” Social Network Analysis and Mining, vol. 14, Art. no. 223, 2024, doi: 10.1007/s13278-024-01380-0.
30. L. S. Abdulzahra and A. J. Obaid, “Detection of fake reviews in Yelp dataset using machine learning and chain classifier approach,” in Micro-Electronics and Telecommunication Engineering, D. K. Sharma, S.-L. Peng, R. Sharma, and G. Jeon, Eds. Singapore: Springer, 2024, pp. 331–346, doi: 10.1007/978-981-99-9562-2_27.

31. C. R. Veda, N. Apoorva, N. Vishnu, M. S. Velpuru, H. Akshaya, and S. P. Siripuram, “Fake review identification using hybrid fusion of machine learning and natural language processing techniques,” in Proc. 3rd Odisha Int. Conf. Electrical Power Engineering, Communication and Computing Technology (ODICON), Bhubaneswar, India, 2024, pp. 1–6, doi: 10.1109/ODICON62106.2024.10797635.
32. R. Das et al., “Towards the development of an explainable e-commerce fake review index: An attribute analytics approach,” European Journal of Operational Research, vol. 317, no. 2, pp. 382–400, 2024, doi: 10.1016/j.ejor.2024.03.008.
33. S. Janokar, K. More, R. Raikwar, K. Joshi, P. Pawar, and T. Singh, “Fake review detection system: A machine learning approach with linguistic and semantic analysis,” in Proc. Int. Conf. Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA), Nagpur, India, 2024, pp. 1–6, doi: 10.1109/ICAIQSA64000.2024.10882276.
34. R. Mohawesh et al., “Fake review detection using transformer-based enhanced LSTM and RoBERTa,” International Journal of Cognitive Computing in Engineering, vol. 5, pp. 250–258, 2024, doi: 10.1016/j.ijcce.2024.06.001.
35. J. Liu et al., “A study on fake review detection based on RoBERTa and behavioral features,” Procedia Computer Science, vol. 242, pp. 1323–1330, 2024, doi: 10.1016/j.procs.2024.08.131.
36. A. H. Alshehri, “An online fake review detection approach using famous machine learning algorithms,” Computers, Materials & Continua, vol. 78, no. 2, pp. 2767–2786, 2024, doi: 10.32604/cmc.2023.046838.
37. A. Thilagavathy, P. R. Therasa, J. J. Jasmine, M. Sneha, R. Shree Lakshmi, and S. Yuvanthika, “Fake product review detection and elimination using opinion mining,” in Proc. World Conf. Communication and Computing (WCONF), Raipur, India, 2023, pp. 1–5, doi: 10.1109/WCONF58270.2023.10234996.
38. P. Naresh, S. V. N. P. Pavan, A. R. Mohammed, N. Chanti, and M. Tharun, “Comparative study of machine learning algorithms for fake review detection with emphasis on SVM,” in Proc. Int. Conf. Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 170–176, doi: 10.1109/ICSCSS57650.2023.10169190.
39. S. Zabeen, A. Hasan, M. F. Islam, M. S. Hossain, and A. A. Rasel, “Robust fake review detection using uncertainty-aware LSTM and BERT,” in Proc. IEEE 15th Int. Conf. Computational Intelligence and Communication Networks (CICN), Bangkok, Thailand, 2023, pp. 786–791, doi: 10.1109/CICN59264.2023.10402342.
40. K. Pal, S. Poddar, S. L. Jayalakshmi, M. Choudhury, Ahmed, and S. Halder, “Opinion mining-based fake review detection using deep learning technique,” in Lecture Notes in Electrical Engineering, 2023, pp. 13–20, doi: 10.1007/978-981-99-2058-7_2.
41. P. Wang, Y. Lin, and J. Chai, “Unmasking deception: A comparative study of tree-based and transformer-based models for fake review detection on Yelp,” in Proc. IEEE Int. Conf. Systems, Man, and Cybernetics (SMC), Honolulu, HI, USA, 2023, pp. 1848–1853, doi: 10.1109/SMC53992.2023.10394573.
42. J. Lu, X. Zhan, G. Liu, X. Zhan, and X. Deng, “BSTC: A fake review detection model based on a pre-trained language model and convolutional neural network,” Electronics, vol. 12, no. 10, Art. no. 2165, 2023, doi: 10.3390/electronics12102165.
43. D. Zhang, W. Li, B. Niu, and C. Wu, “A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information,” Decision Support Systems, vol. 166, Art. no. 113911, Mar. 2023, doi: 10.1016/j.dss.2022.113911.
44. D. U. Vidanagama et al., “Ontology based sentiment analysis for fake review detection,” Expert Systems with Applications, vol. 206, Art. no. 117869, 2022, doi: 10.1016/j.eswa.2022.117869.
45. G. S. Budhi et al., “Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews,” Electronic Commerce Research and Applications, vol. 47, Art. no. 101048, 2021, doi: 10.1016/j.elerap.2021.101048.