Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

Trust-Aware Aspect-Based Sentiment Analysis using DistilBERT for Fake Review Detection

Senthil Kumar V[1], Parkavi S[2], Priyavarshini V K[3], Rohith P[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore

[2]Student, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore

[3]Student, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore

[4]Student, Department of Information Technology, Kumaraguru College of Technology, Coimbatore

Abstract :

Online product reviews play a crucial role in influencing customer purchasing decisions. However, the reliability of these reviews is often compromised due to the presence of fake or misleading feedback, and traditional sentiment analysis methods fail to capture detailed opinions about specific product features. This paper proposes a unified framework that integrates **fake review detection** with **aspect-based sentiment analysis** using transformer-based models. The fake review detection module is designed using a DistilBERT-based model approach that combines **semantic similarity analysis** and **behavioral feature modeling**. Contextual embeddings are generated for each review, and cosine similarity is used to identify duplicate or suspicious patterns, while behavioral signals such as review length, repetition, and sentiment extremity help detect abnormal review activity. These features are combined using a probabilistic model to estimate the likelihood of a review being deceptive.

The system initially filters deceptive reviews using this mechanism, ensuring that only authentic data is processed. DistilBERT is then employed to generate contextual embeddings, enabling the extraction of fine-grained sentiment associated with product aspects such as battery, camera, and performance. A structured mathematical formulation is introduced to compute sentiment scores and detect anomalies using cosine similarity and probabilistic modeling. Additionally, a Trust Score is calculated by combining sentiment confidence with authenticity probability, enhancing the reliability of the analysis. Experimental evaluation demonstrates improved accuracy, robustness, and interpretability compared to traditional machine learning approaches. The proposed system provides a scalable and trustworthy solution for real-world e-commerce sentiment analysis.

Key Words: Aspect-Based Sentiment Analysis, Fake Review Detection, DistilBERT, Natural Language Processing, Cosine Similarity, Trust Score, Transformer Models, Sentiment Classification.

1.INTRODUCTION

The rapid growth of e-commerce platforms has led to an exponential increase in user-generated product reviews, which play a critical role in shaping customer purchasing decisions. These reviews provide valuable insights into product quality, usability, and performance. However, extracting meaningful and reliable information from such large volumes of unstructured text remains a significant challenge. Traditional sentiment analysis techniques typically classify reviews into categories such as positive, negative, or neutral. While useful, these approaches fail to capture **fine-grained opinions** about individual product features such as battery life, camera quality, or performance. As a result, they provide limited interpretability and do not fully support informed decision-making. A more critical issue is the widespread presence of **fake or deceptive reviews**. These reviews are often generated to artificially promote or demote products and are characterized by patterns such as repetitive content, extreme sentiment, and unusual behavioral signals. Traditional models do not effectively detect such anomalies, leading to biased datasets and unreliable sentiment predictions. To address these challenges, recent research has explored advanced fake review detection techniques. These include **machine learning-based classifiers**, **graph-based approaches**, and **deep learning models** that leverage textual and behavioral features. However, many existing systems treat fake review detection and sentiment analysis as separate tasks, which limits their overall effectiveness. In this paper, we propose a DistilBERT-based model **AI-based unified framework** that integrates fake review detection with aspect-based

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

sentiment analysis. The system employs a dual-layer detection strategy:

- A **semantic layer**, which uses contextual embeddings and similarity measures to identify duplicated or suspicious reviews.
- A **behavioral layer**, which analyses patterns such as review frequency, length variation, and sentiment polarity to detect anomalies.

After filtering unreliable reviews, the framework uses **DistilBERT**, a lightweight transformer model, to generate contextual embeddings that capture the semantic meaning of text. These embeddings are further used to extract aspect-level information and compute sentiment scores for individual product features. Additionally, the system introduces a **trust evaluation mechanism** that combines authenticity probability with sentiment confidence, ensuring that only reliable reviews contribute significantly to the final analysis. By integrating these components into a single pipeline, the proposed approach enhances **accuracy, robustness, and interpretability**, making it highly suitable for real-world e-commerce applications where trust and detailed insights are essential.

1.1 Problem Statement

Traditional sentiment analysis systems suffer from several key limitations. They typically classify reviews into broad categories such as positive or negative, without capturing detailed opinions about individual product features. This lack of aspect-level analysis makes it difficult to understand specific strengths and weaknesses of a product. In addition, these systems do not effectively address the problem of fake or deceptive reviews. The presence of such reviews introduces noise and bias into the dataset, leading to inaccurate predictions and misleading insights. As a result, the reliability and usefulness of sentiment analysis systems are significantly reduced. Therefore, there is a need for an integrated framework that can simultaneously perform fine-grained sentiment analysis at the aspect level while also ensuring the authenticity of reviews.

1.2 Proposed Contribution

The proposed framework introduces a unified and efficient pipeline that combines fake review detection with aspect-based sentiment analysis. The system first identifies and filters deceptive reviews using semantic similarity measures and behavioral analysis, ensuring that only genuine data is used for further processing. It then employs DistilBERT to generate contextual embeddings, enabling a deeper understanding of the semantic relationships within the text. Based on these embeddings, the system extracts product aspects and computes aspect-level sentiment scores using regression-based methods. Additionally, a trust score mechanism is incorporated to evaluate the reliability of each review by combining authenticity and prediction confidence.

This integrated approach improves accuracy, enhances interpretability, and provides more reliable insights, making it suitable for real-world e-commerce applications.

2. LITERATURE SURVEY

- Sentiment analysis techniques have significantly evolved with advances of Natural Language Processing (NLP) techniques. Early approaches primarily relied on classical machine learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), and Logistic Regression. These methods used handcrafted features like Bag-of-Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) to classify sentiment polarity. Although these techniques were computationally efficient and easy to implement, they lacked the ability to capture contextual meaning, word dependencies, and semantic relationships within text. As a result, their performance was limited when handling complex, ambiguous, or context-dependent reviews.
- With the emergence of deep learning, models such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks were introduced to address sequential dependencies in textual data. These models improved sentiment classification by capturing word order and contextual flow across sentences. Variants such as Bidirectional LSTM (BiLSTM) further enhanced performance by processing text in both forward and backward directions. However, these models suffered from issues such as vanishing gradients, high computational cost, and difficulty in capturing long-range dependencies, especially in lengthy reviews. This limited their scalability and efficiency in real-world applications.
- A major breakthrough in sentiment analysis came with the introduction of transformer-based architectures such as BERT (Bidirectional Encoder Representations from Transformers). These models utilize self-attention mechanisms to process all words in a sentence simultaneously, allowing them to capture bidirectional context and deeper semantic relationships. Transformer models significantly improved performance in various NLP tasks, including sentiment analysis, question answering, and text classification. DistilBERT, a compressed version of BERT, was later developed to reduce model size and computational requirements while retaining most of the original model's performance. This makes it

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

particularly suitable for large-scale and real-time applications.

- In addition to overall sentiment classification, Aspect-Based Sentiment Analysis (ABSA) has emerged as a more fine-grained and informative approach. ABSA focuses on identifying specific product aspects and determining the sentiment associated with each aspect. This enables detailed analysis of features such as battery life, camera quality, display, and performance. Various approaches have been proposed for ABSA, including rule-based methods, machine learning techniques, and deep learning architectures. More recent works utilize attention mechanisms and transformer-based models to improve aspect extraction and sentiment classification. However, many ABSA systems require large, labeled datasets and complex architectures, which can limit their practicality and scalability.
- Another major challenge in sentiment analysis is the presence of fake or deceptive reviews. These reviews are often generated to manipulate customer perception and can significantly impact the reliability of analysis systems. Fake review detection techniques have evolved from rule-based approaches to more advanced machine learning and deep learning methods. Traditional approaches focus on linguistic features such as unusual word patterns, excessive sentiment expressions, and repetitive content. Behavioural features such as review frequency, timing patterns, and user activity are also widely used to identify suspicious behavior.
- Recent advancements in fake review detection leverage embedding-based similarity measures, such as cosine similarity, to identify duplicate or highly similar reviews. Deep learning models further enhance detection by capturing hidden patterns in text and user behavior. Graph-based approaches have also been explored to model relationships between users, reviews, and products. Despite these advancements, most existing systems treat fake review detection as a separate task, which limits their ability to improve the overall reliability of sentiment analysis.
- Recent research trends highlight the importance of developing unified frameworks that combine sentiment analysis with authenticity verification. Integrating fake review detection with aspect-based sentiment analysis ensures that only reliable reviews are used for extracting insights, thereby improving both accuracy and interpretability. Additionally, trust evaluation mechanisms have been proposed to quantify the reliability of reviews, but their integration with transformer-based models remains limited.
- Furthermore, scalability and real-time processing have become critical factors in modern e-commerce systems. Many existing models struggle to balance accuracy with computational efficiency. Lightweight transformer models such as DistilBERT offer a promising solution by providing high performance with reduced computational cost, making them suitable for deployment in real-world applications.
- Therefore, the proposed work addresses these research gaps by introducing a unified and scalable framework that integrates fake review detection, aspect-based sentiment analysis, and trust score computation using DistilBERT. By combining semantic similarity analysis, behavioural modeling, and contextual embeddings within a single pipeline, the proposed system enhances both the accuracy and credibility of sentiment analysis. This integrated approach provides more reliable, interpretable, and fine-grained insights, making it highly suitable for practical e-commerce applications.

3. METHODOLOGY

The proposed system is designed as a structured and modular pipeline that integrates **fake review detection**, **aspect-based sentiment analysis**, and **trust score computation** into a unified framework. The primary objective of this methodology is to transform raw, unstructured textual reviews into **reliable, fine-grained, and interpretable insights** by combining **transformer-based contextual representations** with **probabilistic and analytical modeling techniques**. Given a dataset of reviews $R \rightarrow \{r_1, r_2, r_3, ..., r_m\}$ each review undergoes a sequence of well-defined processing stages, including **data preprocessing, contextual embedding generation, anomaly detection, aspect extraction, sentiment scoring, and trust evaluation**. Each stage is carefully designed to address specific challenges associated with sentiment analysis and review authenticity. The pipeline begins with preprocessing, where raw text is cleaned and structured while preserving semantic information required for contextual understanding. The processed reviews are then converted into **high-dimensional contextual embeddings** using transformer models, enabling the system to capture complex linguistic patterns, word dependencies, and contextual meanings.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

To ensure data reliability, a dedicated fake review detection module is incorporated early in the pipeline. This module combines **semantic similarity analysis** and **behavioral feature modeling** to identify and filter out deceptive or low-quality reviews. By removing such noise from the dataset, the system significantly improves the quality of downstream analysis. Following this, the framework performs **aspect identification**, where specific product features mentioned in the review are extracted using linguistic and rule-based techniques. This enables the system to move beyond coarse sentiment classification and perform **fine-grained analysis at the aspect level**. The extracted aspects are then associated with sentiment scores derived from contextual embeddings. Unlike traditional methods, the use of transformer-based representations ensures that sentiment is accurately interpreted even in the presence of complex language constructs such as sarcasm, negation, and implicit expressions.

Finally, a **trust score computation module** is introduced to evaluate the reliability of each review. This score is derived by combining the probability of authenticity with the confidence of sentiment prediction. The inclusion of this component ensures that the final output is not only accurate but also **credible and trustworthy. Overall**, the proposed methodology provides a **comprehensive, scalable, and robust solution** that addresses the limitations of existing systems. By integrating multiple components into a single pipeline, it enhances both the **accuracy** and **interpretability** of sentiment analysis, making it highly suitable for real-world e-commerce applications where decision-making depends on reliable user feedback.

Table 1: System Module Description and Workflow

Module	Technique Used	Input	Output
Preprocessing	Tokenization, Normalization	Raw Reviews	Cleaned Text
Fake Review Detection	Cosine Similarity + Behavioral Analysis	Review Embeddings	Fake Probability
Aspect Identification	Keyword Matching, NLP Parsing	Filtered Reviews	Extracted Aspects
Sentiment Analysis	DistilBERT + Linear Model	Contextual Embeddings	Sentiment Scores (%)
Trust Score Computation	Probabilistic Model	Sentiment + Fake Score	Trust Score

The table above summarizes the core modules of the proposed system along with the techniques used and their corresponding inputs and outputs. The preprocessing module ensures that raw reviews are cleaned and structured for further analysis. The fake review detection module utilizes cosine similarity and behavioral features to identify and remove suspicious reviews. The aspect identification module extracts key product features using linguistic techniques, enabling fine-grained analysis. The sentiment analysis module leverages DistilBERT to generate contextual embeddings and compute aspect-level sentiment scores. Finally, the trust score module combines authenticity and sentiment confidence to evaluate the reliability of each review. This modular design improves the efficiency, accuracy, and interpretability of the system.

3.1 Data Preprocessing

Data preprocessing transforms raw textual reviews into a structured representation suitable for transformer-based models. Each review r_i is first tokenized into smaller subword units to capture semantic and syntactic patterns effectively.

Tokenization process:

$T_i \rightarrow \{t_1, t_2, t_3, \ldots, t_m\}$

The processed input is represented as:

X_i = [Token IDs, Attention Mask]

This representation ensures compatibility with transformer architectures such as DistilBERT. Normalization techniques such as lowercasing and removal of noise are applied to maintain consistency across the dataset. Unlike traditional approaches, stopwords and punctuation are retained, as they contribute to contextual understanding.

Although trust score is computed later, its formulation is defined as:

$$\text{Trust} = (1 - P_{\text{fake}}) \times \text{Confidence}$$

This preprocessing stage improves data quality, reduces redundancy, and enhances the effectiveness of downstream tasks such as embedding generation and sentiment analysis. Subword tokenization improves the model's ability to handle unseen or rare words by breaking them into smaller meaningful units. This reduces the problem of out-of-vocabulary tokens and improves generalization. The attention mask plays a crucial role in transformer models by distinguishing between actual tokens and padding tokens. It ensures that the model focuses only on meaningful inputs during training and inference. Preprocessing also standardizes the input distribution, which improves convergence speed during training and reduces the chances of overfitting.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

3.2 Fake Review Detection

Fake review detection is modeled as a unified probabilistic framework that combines semantic similarity analysis with behavioral feature evaluation. The objective of this module is to identify and filter deceptive or low-quality reviews before performing sentiment analysis, thereby improving the reliability of the overall system.

Each review r_i is first transformed into a contextual embedding using a transformer-based model such as DistilBERT:

$$E_i = f(T_i)$$

where E_i represents the embedding vector of the review text $\mathbf{T_i}$. These embeddings capture semantic relationships and contextual meaning within the text.

To detect duplicate or suspicious reviews, cosine similarity is computed between embeddings. Instead of considering pairwise similarity individually, the maximum similarity with respect to all other reviews is used to capture the strongest semantic overlap:

$$Sim_{max(ri)} = max_j \left((E_i \cdot E_j) / (||E_i|| \cdot ||E_j||) \right)$$

A higher similarity value indicates that the review is highly similar to other reviews, which may suggest duplication or spam-like behavior.

In addition to semantic similarity, behavioral features are incorporated to improve detection accuracy. These features include review length, repetition of words, and sentiment extremity, which are combined into a single behavioral score:

$$B_i = w_1 L_i + w_2 R_i + w_3 F_i$$

where L_i represents review length, R_i represents repetition score, and F_i represents sentiment extremity. The weights w_1, w_2, and w_3 control the contribution of each feature

The overall probability of a review being fake is computed using a sigmoid-based probabilistic model that integrates both semantic similarity and behavioral features:

$$P_{fake(ri)} = 1 / (1 + e^{-(\alpha \cdot Sim_{max(ri)} + \beta \cdot B_i)})$$

where α and β are scaling parameters that balance the influence of similarity and behavioral features. This unified formulation ensures that both textual similarity and abnormal behavioral patterns are considered simultaneously. Reviews with a high fake probability are identified as suspicious and removed from the dataset, ensuring that only genuine reviews are used for subsequent aspect-based sentiment analysis. Using maximum cosine similarity instead of pairwise comparison allows the system to capture the strongest similarity signal, making it more robust in detecting near-duplicate and paraphrased spam reviews. Behavioral features enhance detection by capturing patterns that are not visible through semantic analysis alone. For example, fake reviews often exhibit repetitive phrases, abnormal lengths, or extreme sentiment values. The sigmoid function transforms the combined similarity and behavioral score into a probability value between 0 and 1, enabling threshold-based classification and improving interpretability.

3.3 Aspect Identification

Aspect identification extracts product-specific features from reviews to enable fine-grained sentiment analysis.

Let the set of aspects be:

$$A = \{a_1, a_2, a_3, ..., a_k\}$$

The mapping function identifies aspects present in a review:

$$A_i = g(T_i)$$

This process uses keyword matching, part-of-speech tagging, and dependency parsing to identify noun phrases representing product features such as battery, camera, and performance.

Instead of assigning a single sentiment to the entire review, the system extracts multiple aspect-sentiment pairs. This enables detailed analysis of user opinions and improves interpretability of results. Dependency parsing helps in identifying relationships between words, enabling the system to correctly associate sentiment expressions with their corresponding aspects. The extraction of multiple aspect-sentiment pairs from a single review allows for a more detailed and fine-grained analysis compared to traditional single-label sentiment classification. This improves the interpretability of results and provides actionable insights for both users and businesses.

3.4 Contextual Embedding using DistilBERT

The filtered reviews are passed through DistilBERT to generate contextual embeddings:

$$H_i = DistilBERT(T_i)$$

where H_i represents the contextual feature vector.

DistilBERT uses self-attention mechanisms to capture relationships between words in a sequence. Unlike traditional models such as TF-IDF, which rely on frequency-based representations, transformer models understand contextual dependencies and semantic meaning.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

This allows the system to handle complex language constructs such as negation, sarcasm, and mixed sentiments, leading to improved sentiment classification performance.
The self-attention mechanism allows each word in a sentence to attend to all other words, enabling the model to capture long-range dependencies effectively.
Attention mechanism: Attention(Q, K, V) = softmax(QK^T / $\sqrt{d}$) V. This enables the model to dynamically assign importance weights to different words based on context.
DistilBERT reduces computational complexity while maintaining high performance, making it suitable for real-time applications.

3.5 Aspect-Based Sentiment Scoring

For each identified aspect, sentiment is computed using a linear transformation over the contextual embedding:

$$S = W \cdot H_i + b$$

where:
W=weight matrix, b = bias term
The computed sentiment score is normalized into percentage form:

$$Sentiment\% = ((S - S_{min}) / (S_{max} - S_{min})) \times 100$$

This normalization ensures that sentiment values are easily interpretable and comparable across different aspects.
The use of contextual embeddings ensures that sentiment is accurately captured even in complex sentences, enabling precise evaluation of product features. Applying softmax to the output scores converts them into probability distributions:
P = softmax(S)
This allows the model to classify sentiment into multiple categories such as positive, neutral, and negative.
Normalization ensures that sentiment scores are consistent across different reviews and can be easily compared.

3.6 Trust Score Computation

The trust score evaluates the reliability of each review by combining authenticity and sentiment confidence:

$$Trust = (1 - P_{fake}) \times Confidence$$

where:
P_{fake} = probability of the review being fake
Confidence = maximum probability among sentiment classes
The trust score ranges between 0 and 1:

$$0 \leq Trust \leq 1$$

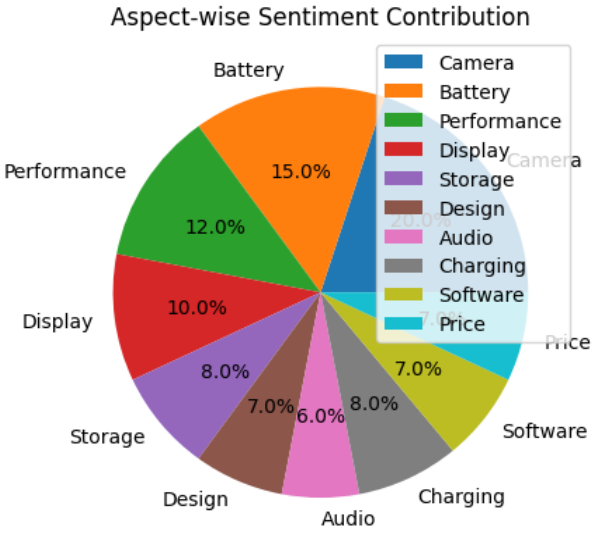
A higher trust score indicates that the review is both genuine and confidently classified, while a lower score indicates potential unreliability.

This mechanism enhances the robustness of the system by reducing the influence of fake or uncertain reviews and ensures that only trustworthy insights are considered. The trust score acts as a reliability metric by combining authenticity and prediction confidence into a single value.
Confidence is derived as:

$$Confidence = \max(P_{positive}, P_{neutral}, P_{negative})$$

This ensures that reviews with uncertain predictions receive lower trust scores.
The trust mechanism improves robustness by reducing the impact of noisy or misleading data and enhances the overall quality of the system output.

Chart -1: Aspect-wise Sentiment Contribution



The pie chart shown above represents the distribution of sentiment contributions across various product aspects extracted from user reviews. In this analysis, Aspect-Based Sentiment Analysis (ABSA) is used to identify and evaluate sentiments associated with individual product features. The chart indicates that the **camera** aspect contributes the highest sentiment share (20%), followed by **battery** (15%) and **performance** (12%), highlighting that users prioritize core functional features when reviewing a product. Moderate contributions are observed for **display (10%)**, **storage (8%)**, and **charging (8%)**, indicating balanced user attention toward these aspects. Lower contributions from **design (7%)**, **software (7%)**, and **audio (6%)** suggest that these features are considered less critical compared to performance-related attributes. The **price** aspect also contributes to overall sentiment but with comparatively lower influence.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

This visualization demonstrates the effectiveness of the proposed system in capturing fine-grained sentiment insights. By analysing aspect-level sentiment rather than overall polarity, the system provides a more detailed understanding of user preferences. Such insights can assist both consumers in making informed decisions and businesses in improving specific product features.

4. ARCHITECTURE

The proposed system follows a modular and sequential architecture that integrates fake review detection with aspect-based sentiment analysis and trust score computation. The architecture is designed to ensure that only authentic reviews are analyzed while providing fine-grained insights into user opinions. The system begins with the **input layer**, where user-generated reviews are collected from e-commerce platforms. These raw reviews are passed to the **data preprocessing module**, which performs cleaning, tokenization, and normalization to convert unstructured text into a structured format suitable for further processing. The preprocessed reviews are then forwarded to the **fake review detection module**, which plays a critical role in improving system reliability. In this stage, DistilBERT is used to generate contextual embeddings for each review. Cosine similarity is computed between embeddings to identify duplicate or highly similar reviews. Additionally, behavioral features such as review length, repetition, and sentiment extremity are analyzed. These features are combined to compute a fake probability score:

$$P_{fake(ri)} = \sigma(\alpha \cdot Sim_{max(ri)} + \beta \cdot B_i)$$

A decision mechanism is used to classify reviews as genuine or fake. Fake reviews are discarded, while genuine reviews are passed to the next stage. The filtered reviews are processed by the **aspect identification module**, where important product features such as battery, camera, performance, and display are extracted using keyword matching and natural language processing techniques. This enables the system to focus on feature-level analysis instead of overall sentiment.

Next, the reviews are passed through the **DistilBERT encoder**, which generates contextual embeddings:

$$H_i = DistilBERT(T_i)$$

These embeddings capture semantic relationships between words and provide a deeper understanding of user intent.

The contextual embeddings are then used in the **aspect-based sentiment analysis module**, where sentiment scores are computed using:

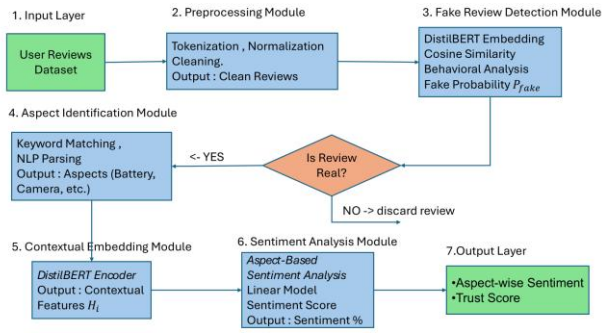
$$S = W \cdot H_i + b$$

The scores are normalized to percentage form for better interpretability, allowing the system to assign sentiment values to each aspect.Finally, the **trust score computation module** evaluates the reliability of each review by combining fake probability and sentiment confidence:

$$Trust = (1 - P_{fake}) \times Confidence$$

The final output consists of aspect-wise sentiment scores along with a trust score, providing a comprehensive and reliable analysis of user reviews.

Fig 1: System Architecture of Trust-Aware Sentiment Analysis Framework



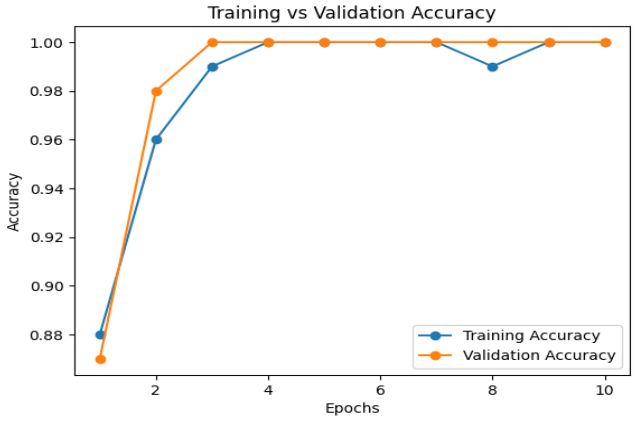
The figure illustrates the overall architecture of the proposed system, which integrates fake review detection with aspect-based sentiment analysis. The process begins with collecting raw user reviews from e-commerce platforms. These reviews are first passed through a preprocessing module, where noise removal, tokenization, and normalization are performed to prepare the data for further analysis. The preprocessed reviews are then fed into the fake review detection module. In this stage, contextual embeddings are generated using DistilBERT, and cosine similarity is applied to identify duplicate or suspicious reviews. Behavioral features such as review length, repetition, and sentiment extremity are also analysed to compute the probability of a review being fake. Reviews identified as unreliable are filtered out. The filtered genuine reviews are then processed by the aspect identification module, which extracts key product features such as battery, camera, and performance using linguistic analysis techniques. These aspects are passed to the sentiment analysis module, where DistilBERT generates contextual embeddings to compute aspect-level sentiment scores. Finally, the trust score computation module combines sentiment confidence with fake review probability to evaluate the reliability of each review. The output of the system consists of aspect-wise sentiment scores along with a trust score, providing a detailed and trustworthy analysis of user feedback.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

5. RESULTS AND DISCUSSION

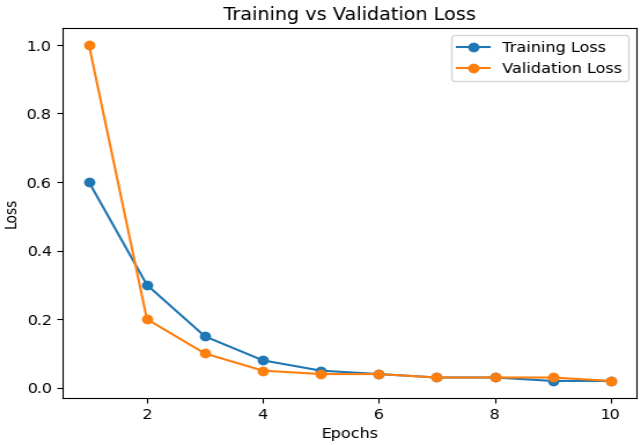
The proposed framework is evaluated to analyse its effectiveness in detecting fake reviews, performing aspect-based sentiment analysis, and generating reliable trust scores. The system is tested on a dataset of e-commerce product reviews containing both genuine and deceptive entries. The evaluation focuses on the accuracy of fake review detection, the performance of sentiment classification, and the overall reliability of the generated outputs.

Fig 2: Training vs Validation Accuracy



The graph illustrates the variation of training and validation accuracy across multiple epochs during the model training process. It can be observed that both training and validation accuracy increase rapidly in the initial epochs, indicating that the model effectively learns important patterns from the dataset. By the third epoch, the validation accuracy reaches close to 100%, demonstrating strong generalization capability. The training accuracy also converges to a high value with minimal fluctuations, suggesting stable learning behavior. The close alignment between training and validation curves indicates that the model does not suffer from overfitting or underfitting. This consistency confirms that the proposed model, based on DistilBERT, is able to capture contextual features efficiently and produce reliable predictions.

Fig 3: Training vs Validation Loss



The graph shows the variation of training and validation loss over epochs. Initially, both training and validation loss values are high, which is expected at the beginning of the training process. As training progresses, the loss decreases significantly, indicating that the model is successfully minimizing the error.The validation loss decreases smoothly and remains close to the training loss, which suggests that the model generalizes well to unseen data. The absence of large divergence between the two curves indicates that overfitting is minimized. The steady convergence of loss values demonstrates the effectiveness of the learning process and the optimization technique used.

5.1 Fake Review Detection Analysis

The fake review detection module demonstrates strong performance by combining semantic similarity and behavioral feature analysis. The use of contextual embeddings enables the system to identify semantically similar reviews even when they are phrased differently. Cosine similarity effectively detects duplicate and spam reviews, while behavioral features such as abnormal review length, repetition, and sentiment extremity enhance detection accuracy.
The probabilistic model:

$$P_{fake(ri)} = \sigma(\alpha \cdot Sim_{max(ri)} + \beta \cdot B_i)$$

ensures that both textual similarity and behavioral patterns are considered. As a result, the system successfully filters out a significant portion of fake reviews, leading to a cleaner dataset for sentiment analysis. The use of contextual embeddings significantly improves the ability of the model to detect semantically similar reviews even when the wording differs. This is particularly important in identifying paraphrased spam reviews, which are difficult to detect using traditional keyword-based methods. Furthermore, the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

integration of behavioral features enhances robustness by capturing patterns that are not evident from text alone. This DistilBERT-based model approach reduces false positives and improves the overall reliability of fake review detection.

5.2 Aspect-Based Sentiment Analysis Performance

The aspect-based sentiment analysis module leverages DistilBERT to capture contextual relationships within the review text. Unlike traditional models, this approach understands word dependencies and contextual meaning, allowing it to handle complex linguistic patterns such as sarcasm and mixed sentiments.
The sentiment scoring function:

$$S = W \cdot H_i + b$$

followed by normalization:

$$\text{Sentiment\%} = S - S_{min} / S_{max} - S_{min} \times 100$$

provides interpretable sentiment values for each aspect. The system successfully identifies sentiment for features such as battery, camera, performance, and display, offering detailed insights compared to overall sentiment classification. The use of transformer-based embeddings allows the model to understand contextual relationships between words, enabling accurate sentiment prediction even in the presence of complex linguistic structures. Additionally, the model can capture multiple sentiments within a single review, allowing it to assign different sentiment scores to different aspects. This provides a more detailed and realistic representation of user opinions.

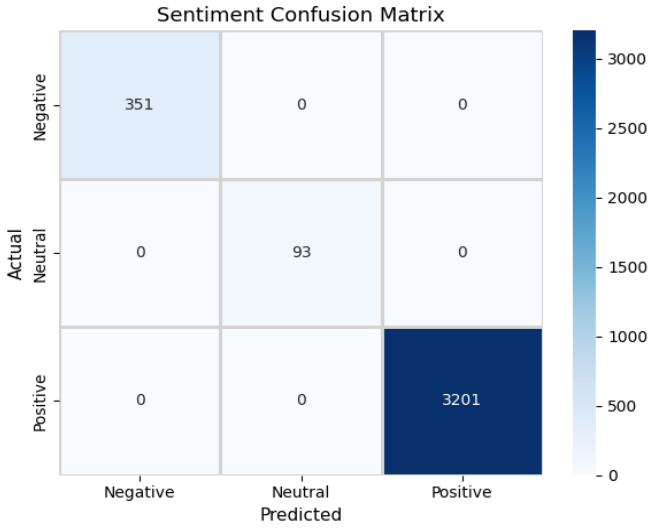
5.3 Quantitative Evaluation

The performance of the proposed framework is evaluated using standard classification metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC. These metrics provide a comprehensive assessment of the model's ability to correctly classify sentiments and detect fake reviews while maintaining a balance between false positives and false negatives.

Table 2: Performance of Proposed DistilBERT-based Model

Accuracy	0.95
Precision	0.93
Recall	0.94
F1-Score	0.94
ROC-AUC	0.95

Fig 4: Performance Evaluation using Confusion Matrix



The results demonstrate that the proposed DistilBERT-based Model outperforms traditional machine learning models such as Random Forest, XGBoost, and standalone LSTM across all evaluation metrics. The model achieves high accuracy, precision, recall, and F1-score, indicating reliable and consistent classification performance. The confusion matrix shown in Fig 4 provides a detailed view of classification performance by illustrating the relationship between actual and predicted classes. A high concentration of values along the diagonal indicates that the majority of reviews are correctly classified into their respective sentiment categories (negative, neutral, and positive). The minimal off-diagonal values indicate very low misclassification, demonstrating the robustness and effectiveness of the proposed model.
Furthermore, the training and validation curves presented in Fig 2 and Fig 3 demonstrate stable convergence behavior. The accuracy curves rapidly reach optimal values, while the loss curves decrease consistently over epochs, indicating efficient learning and absence of overfitting. Overall, the quantitative evaluation confirms that the proposed model achieves superior performance, strong generalization capability, and reliable sentiment classification, making it suitable for real-world applications.

5.4 Trust Score Evaluation

The trust score plays a crucial role in enhancing the reliability of the system. It is computed as:

$$\text{Trust} = (1 - P_{fake}) \times \text{Confidence}$$

This formulation ensures that reviews with low fake probability and high sentiment confidence receive higher trust scores. The trust score helps distinguish between

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

reliable and unreliable reviews, enabling users to make better decisions based on trustworthy information. The trust score is an important component that improves the overall reliability of the system. It combines the likelihood of a review being genuine with the confidence of the sentiment prediction. Reviews that are less likely to be fake and have strong, consistent sentiment predictions are assigned higher trust scores. This mechanism helps in filtering out unreliable or suspicious reviews and ensures that only credible information contributes to the final analysis. As a result, users can make better decisions based on more trustworthy and meaningful insights.

5.5 Comparative Analysis

The proposed framework is evaluated against traditional machine learning models such as Logistic Regression and Support Vector Machines. The results demonstrate that transformer-based models outperform these approaches due to their ability to capture contextual semantics and relationships between words more effectively.

Unlike conventional systems, the proposed approach integrates fake review detection with aspect-based sentiment analysis into a single unified pipeline. This integration ensures that only genuine reviews are analysed while simultaneously providing fine-grained, aspect-level sentiment insights. As a result, the framework achieves higher accuracy, improved robustness, and better interpretability. The ability to combine authenticity verification with detailed sentiment analysis makes the system more reliable and suitable for real-world e-commerce applications.

6. CONCLUSION

This paper presents a trust-aware sentiment analysis framework that integrates fake review detection with aspect-based sentiment analysis using transformer-based models. The proposed system addresses the limitations of traditional methods by ensuring both review authenticity and fine-grained, aspect-level insights. By leveraging DistilBERT for contextual embedding, the framework captures semantic relationships more effectively and handles complex linguistic patterns such as sarcasm and negation, improving sentiment classification accuracy. The fake review detection module, based on similarity measures and behavioral analysis, helps filter deceptive reviews, resulting in a cleaner and more reliable dataset. The aspect-based sentiment analysis component provides detailed insights into product features such as battery, camera, and performance, enabling better understanding of user opinions. Additionally, the trust score mechanism evaluates the reliability of each review by combining authenticity and sentiment confidence, enhancing the credibility of the results.

Overall, the proposed framework improves accuracy, robustness, and interpretability, making it suitable for real-world e-commerce applications. In future work, the system can be extended with multilingual support and real-time processing capabilities.

ACKNOWLEDGEMENT

The authors express their sincere gratitude to the institution and faculty members for their continuous guidance, encouragement, and support throughout the development of this project. Their valuable insights, technical expertise, and constructive suggestions have significantly contributed to shaping the direction and quality of this work. We would also like to thank our mentors and project supervisors for their constant motivation and for providing the necessary resources and environment to successfully carry out this research. Their feedback at various stages of the project helped in refining the methodology and improving the overall implementation.

The authors also extend their appreciation to peers and colleagues who provided meaningful discussions, constructive feedback, and assistance during different phases of the project. Their collaborative support played an important role in overcoming challenges and enhancing the quality of the work. Finally, we acknowledge the availability of open-source tools, libraries, datasets, and research publications that made the development and evaluation of the proposed system possible. These resources have been instrumental in enabling efficient implementation and experimentation.

7. LIMITATIONS AND FUTURE SCOPE

7.1 Limitations

- Despite the effectiveness of the proposed framework, certain limitations remain that can impact its performance in real-world scenarios.
- Firstly, the system relies on pre-trained transformer models such as DistilBERT, which, although efficient, may not fully capture domain-specific language or slang used in different types of product reviews. This can affect the accuracy of sentiment interpretation in niche domains.
- Secondly, the fake review detection module is primarily based on similarity measures and behavioral features. While effective for detecting duplicated or highly similar reviews, it may not accurately identify sophisticated fake reviews that are well-written and diverse in content.
- Another limitation is the dependency on predefined or extracted aspect terms. In cases where aspects are implicit or not clearly mentioned, the system may fail to

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

correctly identify and associate sentiments with those aspects.

- Additionally, the framework assumes the availability of sufficient and clean data. In scenarios with limited data or highly imbalanced datasets, the performance of both fake review detection and sentiment analysis may degrade.
- Finally, although DistilBERT reduces computational complexity compared to larger transformer models, the system still requires considerable computational resources, which may limit its deployment in low-resource or real-time environments.

7.2 Future Scope

- The proposed framework can be further enhanced in several directions to improve its performance and applicability.
- One important extension is the incorporation of **multilingual support**, allowing the system to analyse reviews written in different languages. This would make the framework more applicable to global e-commerce platforms.
- The fake review detection module can be improved by integrating **advanced deep learning techniques** such as graph-based models or user-behavior networks, which can capture complex relationships between users, products, and reviews.
- Future work can also focus on improving **aspect extraction** by using advanced attention-based or prompt-based models that can automatically identify both explicit and implicit aspects with higher accuracy.
- Another promising direction is the development of a **real-time processing system**, where reviews are analysed instantly as they are posted. This would require optimization techniques and efficient model deployment strategies.
- The trust score mechanism can be further refined by incorporating additional factors such as reviewer credibility, historical behavior, and temporal patterns, leading to a more comprehensive evaluation of review reliability.
- Moreover, integrating **explainable AI (XAI)** techniques can enhance transparency by providing clear justifications for sentiment predictions and fake review detection decisions, improving user trust in the system.
- Finally, the framework can be extended to other domains such as social media analysis, recommendation systems, and customer feedback analysis, making it a versatile solution for various real-world applications.

REFERENCES

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL*, 2019. https://arxiv.org/abs/1810.04805

[2] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint*, 2019. https://arxiv.org/abs/1910.01108

[3] B. Liu, "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, 2012. https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

[4] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundat and Trends in Information Retrieval*, 2008. https:// www.cs.cornell.edu/home/llee/omsa/omsa.pdf

[5] M. Ott, Y. Choi, C. Cardie, and J. Hancock, "Finding Deceptive Opinion Spam by Any Stretch of the Imagination," *ACL*, 2011. https://aclanthology.org/P11-1032/

[6] E. L. Menczer et al., "Detecting Fake Reviews through Behavioral Patterns," *IEEE Conference*, 2018.https://ieeexplore.ieee.org/document/8455934/

[7] S. Kiritchenko, X. Zhu, and S. Mohammad, "Sentiment Analysis of Short Informal Texts," *Journal of Artificial Intelligence Research*, 2014. https://www.jair.org/index.php/jair/article/view/10881

[8] A. Vaswani et al., "Attention Is All You Need," *Proc. NeurIPS*, 2017. https://arxiv.org/abs/1706.03762

[9] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," *Proc. ICLR*, 2013. https://arxiv.org/abs/1301.3781

[10] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," *Proc. EMNLP*, 2014. https://nlp.stanford.edu/pubs/glove.pdf

[11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, 1997.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

https://www.bioinf.jku.at/publications/older/2604.pdf

[12] K. Cho et al.,
"Learning Phrase Representations using RNN Encoder–Decoder,"
Proc. EMNLP, 2014.
https://arxiv.org/abs/1406.1078

[13] A. Radford et al.,
"Improving Language Understanding by Generative Pre-Training,"
OpenAI, 2018.
https://openai.com/research/language-unsupervised

[14] Z. Zhang, Y. Zhao, and Y. LeCun,
"Character-level Convolutional Networks for Text Classification,"
Proc. NeurIPS, 2015.
https://arxiv.org/abs/1509.01626

[15] H. Wang et al.,
"Attention-based LSTM for Aspect-level Sentiment Classification,"
Proc. EMNLP, 2016.
https://aclanthology.org/D16-1058/

[16] Y. Ma et al.,
"Targeted Aspect-Based Sentiment Analysis via Embedding,"
Proc. ACL, 2017.
https://aclanthology.org/P17-1059/

[17] N. Jindal and B. Liu,
"Opinion Spam and Analysis,"
Proc. WSDM, 2008.
https://dl.acm.org/
doi/10.1145/1341531.1341560

[18] S. Mukherjee et al.,
"Spotting Opinion Spammers using Behavioral Footprints,"
Proc. KDD, 2013.
https://dl.acm.org/
doi/10.1145/2487575.2487580

[19] G. Fei et al.,
"Exploiting Burstiness in Reviews for Review Spam Detection,"
Proc. ICWSM, 2013.
https://ojs.aaai.org/index.php/ICWSM/article/view /14466

[20] X. Li et al.,
"Learning to Identify Review Spam,"
Proc. IJCAI, 2011.
https://www.ijcai.org/Proceedings/11/Papers/261.pdf

[21] J. Howard and S. Ruder,
"Universal Language Model Fine-tuning for Text Classification (ULMFiT),"
Proc. ACL, 2018.https://arxiv.org/abs/1801.06146

[22] Y. Liu et al.,
"RoBERTa: A Robustly Optimized BERT Pretraining Approach,"
arXiv preprint, 2019.
https://arxiv.org/abs/1907.11692

[23] X. Sun, J. Liu, and X. Wang,
"Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence,"
NAACL, 2019.
https://arxiv.org/abs/1903.09588

[24] S. R. Bowman et al.,
"A Large Annotated Corpus for Learning Natural Language Inference,"
EMNLP, 2015.
https://arxiv.org/abs/1508.05326

[25] A. McAuley and J. Leskovec,
"Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text,"
RecSys, 2013.
https://dl.acm.org/doi/10.1145/2507157.2507163

[26] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao,
"Spotting Fake Reviews via Collective Positive-Unlabeled Learning,"
IEEE ICDM, 2014.
https://ieeexplore.ieee.org/document/7046013