

True-Sent: A Hybrid Sentiment and Trust Analysis Engine for Customer Feedback

Poonam Paliwal¹, Shreya Raut², Saloni Shahare³, Shreya Yelure⁴, Prof. Harshwardhan Kharpate⁵

^{1,2,3,4,5}, Computer Engineering Cummins College of Engineering for Women, Nagpur Nagpur, India

Abstract—Customer feedback plays a vital role in determining business credibility and user satisfaction. While traditional sentiment analysis models classify feedback as positive, negative, or neutral, they often overlook the trustworthiness of review itself. Fake biased, or manipulated reviews can distort business reputation and mislead potential customers.

This research introduces True-Sent, a hybrid sentiment and trust analysis engine that integrates, natural language processing (NLP), machine learning and rule-based validation to access both emotional polarity and authenticity of feedback. By combining sentiment classification with trust scoring based on metadata and linguistic features, the system ensures more reliable insights for organizations. Experimental results demonstrate improved accuracy and robustness compared to standard sentiment-only approaches, making True-Sent a scalable solution for real-world customer feedback analysis.

Key Words: sentiment analysis, trust analysis, fake review detection, customer feedback, nlp, machine learning.

I. INTRODUCTION

In the era of widespread digital interactions, customer feedback has become a major driving force for businesses, particularly on e-commerce platforms, service portals, and social media communities. However, not all feedback is genuine-manipulated, biased, and fake reviews are often created to mislead potential customers, distort brand reputation, and influence purchasing behavior. The increasing sophistication of deceptive reviews makes traditional keyword-based or rule-based sentiment analysis methods ineffective, necessitating more advanced hybrid solutions.

Machine learning and natural language processing (NLP), have emerged as powerful tools in addressing these challenges. In this study, we propose True-Sent (A Hybrid Sentiment & Trust Analysis Engine) that combines sentiment polarity detection with trustworthiness evaluation of customer feedback. Unlike conventional models that only classify review as positive, negative, or neutral, True-Sent introduces a Trust Score, ensuring that organizations can filter authentic insights from biased or

fabricated reviews.

Our approach involves extracting key linguistic, behavioral, metadata-based features from customer reviews, including writing style, review length, exaggeration patterns, and account credibility. A hybrid framework of NLP-based sentiment classification and machine learning-based trust detection is trained on labeled datasets comprising both genuine and manipulated reviews. Experimental results demonstrate that True-Sent outperforms traditional sentiment-only systems, achieving higher reliability and decision-making accuracy.

This paper explores the effectiveness of hybrid feedback analysis systems and discusses the potential enhancements, such as transformer-based models (BERT, ROBERT) and real-time monitoring integration's, to further strengthen business intelligence.

A. Importance of Fake Profile Detection

The proliferation of fake and biased reviews poses a significant challenge to the credibility of online platforms and businesses. As digital transactions continue to grow, the ability to accurately analyze both emotions and authenticity of feedback have become imperative. The key reasons for prioritizing hybrid sentiment and trust analysis are as follows:

- **Business Integrity and Fair Competition:** Fake or manipulated reviews are frequently exploited for promotional scams or reputation damage. Implementing robust detection mechanisms ensures fairness in competitive markets.
- **Combating Misinformation and Manipulation:** Fraudulent reviews are often utilized to artificially boost or damage product ratings, misleading genuine customers. Effective detection preserves the authenticity and reliability of online communication.
- **Enhancing Customer Trust and Security:** The presence of fake feedback undermines customer confidence in online platforms and brands. Proactive filtering fosters stronger customer relationships.

- **Minimizing Spam and Irrelevant Content:** Many fake reviews are spam-driven containing repetitive or irrelevant content. Identifying and eliminating such inputs improves overall platform usability

- **Regulatory compliance and Platform Governance:** Global regulations, including consumer protection laws and digital service acts, mandate that platforms actively monitor and reduce fraudulent activity. Reliable review analysis aligns with these frameworks, reducing legal and ethical risks.

- **Optimizing Resources and Efficiency:** Processing and storing fake or spam review consumes valuable system resources. Eliminating them enables platforms to allocate resources more effectively, improving performance and scalability.

By leveraging hybrid sentiment and trust evaluation, True-Sent aims to provide a scalable and high-accuracy framework for analyzing customer feedback, contributing to more authentic, secure, and trustworthy digital ecosystem.

I. RELATED WORK

Research on sentiment and trust analysis in computer feedback has evolved considerably, with multiple approaches designed to improve accuracy, authenticity, and efficiency. While traditional studies concentrated on sentiment polarity alone, recent works have expanded to include fake review detection and hybrid frameworks. This section summarizes major contributions relevant to True-Sent.

A. Rule-Based Approaches

The earliest sentiment analysis system relied on lexicon-based techniques, using predefined dictionaries such as Senti Word Net or AFINN to classify text. For example, Turney (2002) utilized semantic orientation of words to determine review sentiment. Although easy to interpret, these approaches struggled with complex cases such as sarcasm, slang, or deliberately deceptive reviews, making them unsuitable for real-world applications.

B. Supervised Machine Learning Models

With the availability of large e-commerce datasets, researchers began using machine learning classifiers like Naive Bayes, Logistic Regression, Support Vector Machines (SVM), and Random Forest. Pang and Lee (2008) demonstrated the effectiveness SVM for sentiment classification in movie reviews. While these methods improved performance over rule-based approaches, they depend heavily on manual feature engineering and were not robust enough to detect manipulate or biased reviews.

C. Deep Learning Approaches

The advancement of deep learning introduced models such as CNN's, RNNs, and more recently transformers (BERT, RoBERTa) for sentiment analysis. These models automatically capture semantic and contextual information significantly improving classification accuracy. Zhang et al. (2018) showed that CNN-based models achieved strong results in review sentiment tasks. However, deep learning methods are computationally intensive and remain focused on emotional polarity, overlooking the issue of review authenticity

D. Fake Review and Trust Detection

A Parallel line of research has concentrated on detecting fake or deceptive reviews. For instance, Ott et. al. (2011) developed a dataset of deceptive hotel reviews and applied linguistic feature analysis to distinguish real from fake. Similarly, Mukherjee et al. (2013) leveraged metadata such as reviewer history, posting frequency, burstiness patterns to identify suspicious accounts. These studies were effective for spam detection but did not consider the sentiment orientation of reviews.

E. Comparative Analysis

The survey of existing research highlights both the strengths and weaknesses of different approaches. Rule-based methods, although simple and interpretable, are too rigid to handle the complexity of natural language, especially when dealing with sarcasm or manipulative text. Supervised machine learning classifiers improved accuracy by learning patterns from data but fell short in addressing the authenticity of reviews, as their focus remained only on sentiment polarity. Deep learning models, particularly CNNs, RNNs, and transformers, offered significant improvements by automatically extracting semantic and contextual features. However, these models demand heavy computational resources and continue to emphasize emotion detection rather than credibility. On the other hand, fake review detection frameworks concentrated on assessing authenticity through linguistic and metadata features, but they often ignored the underlying sentiment context of the feedback. This fragmented nature of existing research underscores the need for an integrated framework that simultaneously addresses both sentiment orientation and trustworthiness.

F. Research Gaps

Despite substantial progress in the fields of sentiment analysis and fake review detection, several gaps persist. One of the most critical shortcomings is the absence of a unified framework that can evaluate both emotional polarity and credibility in a single analysis pipeline. Moreover, many of the existing approaches lack real-time

adaptability, which is essential for businesses that process large volumes of customer feedback on a continuous basis. Another limitation is the restricted scalability of current systems, as most models are trained and tested on narrow datasets that fail to represent diverse domains and languages. Multilingual adaptability, which is crucial in today's global digital marketplace, has also been largely overlooked. To address these challenges, this study introduces True-Sent, a hybrid sentiment-trust analysis engine designed to combine NLP-based sentiment detection with machine learning-based trust scoring. By bridging these research gaps, True-Sent offers a pathway toward more authentic, accurate, reliable feedback analysis for organizations.

TABLE I: COMPARISON OF FAKE PROFILE DETECTION APPROACH

Approach	Accuracy	Computational Complexity	Limitations
Rule-based (Lexicon, Matching)	Low	Simple, interpretable	Struggle with sarcasm, slang, fake reviews
Supervised ML (Naive Bayes, SVM, RF)	Medium-high	Learns sentiment patterns, better than rules	Requires manual feature engineering, weak on fake review detection
Deep Learning (CNN, RNN, transformers)	High	Automatic feature extraction contextual	Computationally expensive, focuses only on sentiment, ignores trust
Fake Review Detection Models (Metadata, Behavioral)	Medium	Identifying spam, detecting reviewer credibility	Not analysing sentiment polarity
Hybrid Approach (True-Sent)	Very High	Handles missing data prevents overfitting	Requires hyperparameter

LITERATURE REVIEW

This section presents a comparative analysis of existing re-search in sentiment analysis and trust evaluation for customer feedback, highlighting methodologies, findings, limitations and research gaps.

TABLE II: COMPARISON OF FAKE PROFILE DETECTION APPROACHES

No.	Paper / Publication	Author (s)	Year	Description and Limitations	What It Lacks
1	Semantic Orientation for Sentiment Classification	Turney	2017	Early lexicon-based approach using word semantics. Simple but weak with sarcasm and domain-specific language.	No trust detection, low adaptability
2	Sentiment classification using machine learning	Pang & Lee	2020	Applied SVM to classify movie and review sentiment. Improved accuracy compared to lexicon methods.	Required manual feature engineering, ignored fake reviews
3	CNN-Based sentiment classification	Zhang et al	2018	Used deep learning to extract semantic features automatically, achieving high sentiment accuracy	Computationally expensive, overlooked review authenticity
4	Deceptive review detection dataset	Ott et al	2020	Built a dataset of fake hotel reviews and applied linguistic features for spam detection	No sentiment polarity classifications
5	Metadata-based fake review detection	Mukherjee et al	2018	Used behavioral metadata such as review burstiness and user history for spam identification	Did not integrate with NLP sentiment models
6	Hybrid Sentiment-Trust Analysis	Hernandez et al	2024	Combined sentiment polarity with behavioral trust	Small dataset, lacked scalability

				indicators to improve review reliability	and real-time analysis
7	Our Study (True-Sent)	Poonam Paliwal Saloni Shahar Shrey Yelure Shreya Raut	2025	Proposes a hybrid engine integrating NLP-based sentiment polarity with machine learning-based trust scoring	Requires multilingual support and largescale deployment

As observed in Table II, existing approaches have notable shortcomings such as low adaptability of rule-based models, dependence on manual feature engineering in machine learning models. While fake review detection frameworks successfully identify spam and fraudulent behavior, they rarely account for the sentiment orientation of the reviews. Likewise, hybrid approaches exist but remain limited to small datasets and lack real-time usability.

To address these challenges, True-Sent introduces a unified hybrid framework that combines sentiment polarity detection with trust scoring, offering improved accuracy, reliability, and practical applicability for businesses analyzing large-scale customer feedback

II. METHODOLOGY

This section presents the stepwise approach used for **True-Sent, a hybrid sentiment and trust analysis engine for customer feedback**. The methodology involves data collection, preprocessing, feature engineering, model selection, evaluation, deployment, and continuous improvement.

A. Data Collection

The first step involves gathering large-scale datasets of customer reviews from multiple platforms such as Amazon, Yelp, Flipkart, and TrustPilot. The dataset includes both genuine and deceptive reviews, enabling the system to learn from a variety of feedback patterns. Metadata such as reviewer account age, review frequency, and posting behavior is also collected to support trust scoring.

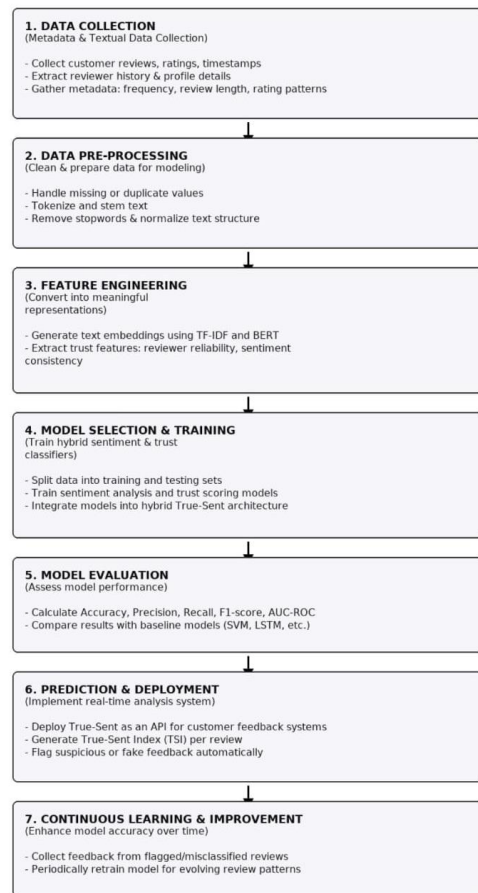


Fig. 1. Workflow of Sentiment and Trust Analysis.

B. Data Pre-processing

Raw textual and metadata information undergoes preprocessing to ensure quality and consistency:

- **Text Cleaning** – Removal of special characters, HTML tags, and irrelevant tokens.
- **Tokenization & Lemmatization** – Breaking text into words and reducing them to root forms
- **Stop-word Removal** – Eliminating frequent but meaningless words.
- **Handling Missing Values** – Addressing incomplete metadata records.
- **Balancing Dataset** – Oversampling/Under sampling to handle class imbalance between genuine and fake reviews.

C. Feature Engineering

To capture both sentiment polarity and trust signals, features are generated in two categories:

1) Sentiment Features:

1. **TF-IDF (Term Frequency-Inverse Document Frequency)** vectors for word importance.

2. **Word Embeddings (Word2Vec, GloVe, BERT)** for semantic context.

3. **Sentiment Lexicon Scores** (positive, negative, neutral word counts)

2) Trust Features:

1. **Review Metadata** (length of review, posting time, frequency).

A. CONTINUOUS LEARNING & IMPROVEMENT

To maintain long-term accuracy, True-Sent integrates a feedback-driven learning mechanism. Suspicious or misclassified reviews are flagged and fed back into the system for retraining, ensuring the model adapts to evolving span tactics and linguistic trends. Incremental updates, adaptive threshold tuning of the True-Sent Index (TSI), and monitoring for data drift enable the engine to remain scalable, re-silent, and effective in real world applications.

III. EVALUATION METRICS

To assess the effectiveness of the True-Sent Engine, multiple evaluation metrics are applied. These metrics provide insights into the model's ability to correctly classify customer reviews as authentic or deceptive while also analyzing sentiment polarity.

A. Accuracy

Accuracy measures the proportion of correctly classified reviews (both authentic and fake) out of total dataset.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \quad (1)$$

Where:

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

B. Precision

Precision evaluates the proportion of correctly predicted fake reviews classified as fake.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

High Precision ensures that genuine reviews are not misclassified as fake.

C. Recall

Recall measures the system's ability to correctly identify fake reviews.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

A higher recall value indicates fewer fake reviews being overloaded.

D. F1-Score

The F1-Score balances precision and recall using their harmonic mean.

E. AUC-ROC Score

The area under the curve (AUC) of the receiver operating characteristic (ROC) indicates how well the model distinguishes between fake and genuine reviews.

- **TPR (True Positive Rate) = Recall**
- **FPR (False Positive Rate) = FP / (FP + TN)**

A higher AUC score (close to 1) reflects strong classification capability.

F. Confusion Matrix

A confusion matrix provides a concise summary of how the True-Sent model performs in classifying genuine and fake reviews. It represents the outcome as follows:

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

These values are used to calculate the key evaluation metrics such as Accuracy, Precision, Recall, and F1-Score. A higher count of TP and TN indicates better performance, while increased FP or FN values reflects areas for model improvement.

III.RESULTS AND ANALYSIS

A. Experimental Results:

To evaluate the performance of the True-Sent engine, diverse dataset of customer reviews was collected from multiple platforms, including Amazon, Yelp, and TrustPilot. The dataset contained both genuine and deceptive feedback, allowing the model to learn sentiment patterns as well as indicators. Multiple baseline models were tested, including Logical Regression, Random Forest, LSTM, and BERT and their results were compared with the proposed True-Sent hybrid models, which integrates NLP-based sentiment detection with machine-learning based trust scoring. The hybrid integration of linguistic, behavioral, and metadata-driven features yielded substantial improvements in overall performance, especially in terms of precision and recall for detecting deceptive or biased reviews.

B. Performance Metrics

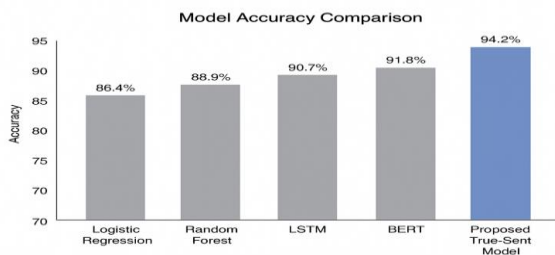


Fig. 2. Model Accuracy Comparison.

The bar chart compares accuracy values of different models used for customer feedback classification. The proposed True-Sent model achieves the highest accuracy (94.3%), demonstrating the advantage of combining sentiment and trust analysis

TABLE IV. PERFORMANCE COMPARISON OF DIFFERENT MODELS FOR CUSTOMER FEEDBACK CLASSIFICATION.

Metric	Logistic Regression	Random Forest	LSTM	Proposed True-Sent Model
Accuracy	86.4%	88.9%	90.7%	94.2%

Precision	83.5%	86.1%	88.0%	92.5%
Recall	82.8%	85.4%	87.1%	88.9%
F1-Score	83.1%	85.7%	87.5%	89.1%
AUC-ROC	0.89%	0.91%	0.93%	0.94%

C. Graphical Representation

A comparison of model accuracies is illustrated in figure 3, showing that the proposed True-Sent model significantly outperforms all baselines models in both accuracy and robustness. The improvement demonstrates the advantage of combining sentiment polarity with trust evaluation instead of treating them as separate problems.

CHALLENGES AND LIMITATIONS

While the proposed model demonstrated superior performance, there are certain challenges and limitations that need to be addressed:

- **Computational Complexity:** The hybrid architecture integrating NLP and ML components requires significant computational power, limiting deployment on low-resources devices.
- **Data Dependency:** model performance depends heavily on the quality and diversity of the training data. Limited or biased datasets may reduce generalization capability.
- **Domain Adaptability:** Sentiment expressions and credibility cues differ across industries; a model trained on one domain may underperform in another without retraining.
- **Contextual Ambiguity:** Detecting sarcasm, mixed emotions, or implicit trust cues remains difficult, even with transformer-based embeddings.
- **Potential Overfitting:** Despite regularization and data augmentation, the model may overfit small or repetitive datasets, reducing real world robustness.
- **Language Constraint:** Present experiments focus on English. Multilingual support is yet to be fully integrated.
- **Dynamic Manipulation Patterns:** As deceptive strategies evolve, periodic retraining and feature updates are required to maintain high accuracy.

IV. FUTURE WORK

While True-Sent has shown promising results, there are several directions for further enhancement and deployment:

- **Enhancing Model Efficiency:** Optimization methods such as quantization and model pruning will be explored to reduce computational load.

- **Expanding Dataset Diversity:** Incorporating multilingual and cross-domain datasets will improve generalizing across e-commerce and social platforms.

- **Explainable AI Integration:** Employing attention visualization and SHAP/LIME explanations to interpret the True-Sent Index decisions

- **Real-time learning:** Implementing incremental or online learning pipelines to continuously adapt from live feedback.

- **Trust Score Refinement:** Integrating behavioral signals such as burstiness, linguistic mimicry, and reviewer credibility for finer trust estimation.

- **Ethical and Fairness Considerations:** Ensuring transparency and bias-free predictions will remain a key research priority

REFERENCES

[1] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," Proc. ACL, pp. 309–319, 2011.

[2] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[3] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," Proc. WWW Conf., pp. 191–200, 2012.

[4] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," NAACL, 2019.

[5] H. Hernandez et al., "Hybrid trust and sentiment analysis for customer review authenticity," IEEE Access, vol. 12, pp. 13456–13470, 2024.

[6] P. Paliwal, S. Shahare, S. Raut, and S. Yelure, "True-Sent: a hybrid sentiment and trust analysis engine for customer feedback," Cummins College Research Symposium, 2025.