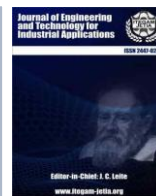




ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.59, p.303-311. May/June, 2026.

DOI: <https://doi.org/10.5935/jetia.v12i59.3435>

RESEARCH ARTICLE

OPEN ACCESS

MEASURING REVIEW CREDIBILITY AND DECEPTION ON E-COMMERCE PLATFORMS THROUGH NLP AND CHAOS THEORY

Preeti Marwaha¹ and Arunita Chaukiyal²

^{1,2}Department of Computer Science, Acharya Narendra Dev College, University of Delhi, New Delhi, India.

¹<https://orcid.org/0009-0000-7425-4749>^{1b}, ²<https://orcid.org/0009-0008-5185-5945>^{2b}

Email: preetimarwaha@andc.du.ac.in, arunita@andc.du.ac.in

ARTICLE INFO

Article History

Received: January 27, 2026

Reviewed: March 28, 2026

Accepted: April 28, 2026

Published: May 29, 2026

Keywords:

NLP,
Sentiment analysis,
Deception,
Trust Score,
E-commerce,
Chaos theory.

ABSTRACT

Customer reviews on e-commerce sites strongly influence customers' choices of products for purchasing but these reviews often include noise, exaggeration, or misleading language that make a product look better or worse than it really is. We propose NETER (Nonlinear Entropy-based Trust Evaluation Rating), a method for evaluating reliability and authenticity of online product reviews by handling non linearities in review text. NETER combines techniques likes sentiment analysis, deception detection, entropy-based stability and chaos-theory modeling to generate a trust score of product. Model uses RoBERTa for sentiment analysis to identify context and hidden emotions in review text accurately. Deception module identifies features like extreme praise, urgency words, or promotional phrases to check reliability of reviews. NETER uses statistical methods on review count to compute confidence weight so that products with few reviews do not receive unfairly high scores. To detect unstable, exaggerated or hyped review patterns, NETER applies chaos-theory where a logistic map is computed to determine Lyapunov exponent. Reviews showing hype patterns, unusual changes in their style or tone receive penalties while consistent sentiment is rewarded. Entropy and dispersion measure sentiments' consistency. Thus, NETER creates a clearer and reliable review score in comparison with sentiments only traditional methods.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Online customer reviews deeply impact how people make buying decisions on e-commerce platforms. Customers trust star ratings, written reviews, and summaries to determine a product's quality, compare options, and decide whether to purchase them or not. But these reviews are not always reliable. Some reviews are excessively praised while some use promotional language, emotional words, or suddenly change emotional tone in successive reviews. All such reviews can be misleading and prompt potential customers to make a favorable decision for the product. Such patterns are often linked to dark pattern design techniques on e-commerce platforms which are majorly employed to influence customers. The traditional sentiment-based methods can overestimate a product's reliability because they don't separate real user experiences from biased or manipulated reviews. To solve this, we propose NETER (Nonlinear Entropy and Trust adjusted Evaluation Rating). NETER is a multi-layered system that combines sentiment analysis, deception detection, statistical reliability, and chaos-theory-based stability modeling. Unlike traditional methods that treat reviews as separate and fixed, NETER understands that review patterns can behave in complex, nonlinear ways. Even small emotional or manipulative cues can create a misleading sense of product quality. The NETER framework begins by identifying the basic sentiment of each review using RoBERTa, a transformer-based NLP model that classifies emotional tone with high accuracy. It then uses a deception-detection module that looks for signs of manipulation, such as promotional wording, urgency phrases, excessive punctuation, and overly dramatic expressions commonly present in low-credibility reviews. To avoid giving unfairly high trust scores to products with only a small number of reviews, NETER uses confidence weight based on review count. The confidence weight adjusts trust score based on how much feedback or customers' review is available. To capture instability which simple statistics cannot detect, we propose the NETER framework which uses a logistic map based on chaos-theory. The model simulates each product's review sentiment and calculates a Lyapunov exponent to measure how sensitive the sentiment is to small changes.

To denote trustworthiness of a product, the NETER penalizes products that show random or erratic or hyped textual reviews. NETER calculates entropy and dispersion to measure consistency in user sentiments. Thus, NETER can help customers to make better choices as it identifies deceptive patterns in text reviews and assigns more reliable scores to the textual reviews of a product.

1.1 PROBLEM STATEMENT

With the absence of a comprehensive, reliable, and stability-aware framework which is based on textual reviews and can measure the reliability of, we need a method that can:

1. Differentiate genuine sentiment from deceptive or promotional language.
2. Adjust trust scores based on review volume and statistical confidence.
3. Detect nonlinear instability in review patterns using chaos-theory principles
4. Integrate these elements into a single, interpretable trust metric.

These issues need to be solved because it makes reviews more trustworthy, helps shoppers choose better, and limits misleading behavior on e-commerce sites.

II. LITERATURE REVIEW

Research in digital spaces now focuses more on user-created content, how people interact with technology, and how algorithms try to influence users' selection and thinking. To address our research goals, we carry out a detailed and well-organized review of the existing literature. This is supplemented by a keyword co-occurrence network analysis to map conceptual relationships and key focus areas within the field. The keyword co-occurrence network as shown in Figure 1, is constructed using the top 25 keywords from the Scopus dataset and visualized using the VOSviewer methodology [1]. In Figure 1, node size represents keyword frequency, and edge thickness represents co-occurrence strength within the same studies. Central clusters revolve around transformer-based NLP (BERT, RoBERTa, DistilBERT), fake review detection and opinion spam, and explainable AI, privacy, and dark patterns. Literature surveys include areas like deceptive and fake review detection, sentiment analysis and rating prediction, dark patterns and manipulative interface design, chaos theory, complex systems, and anomaly detection.

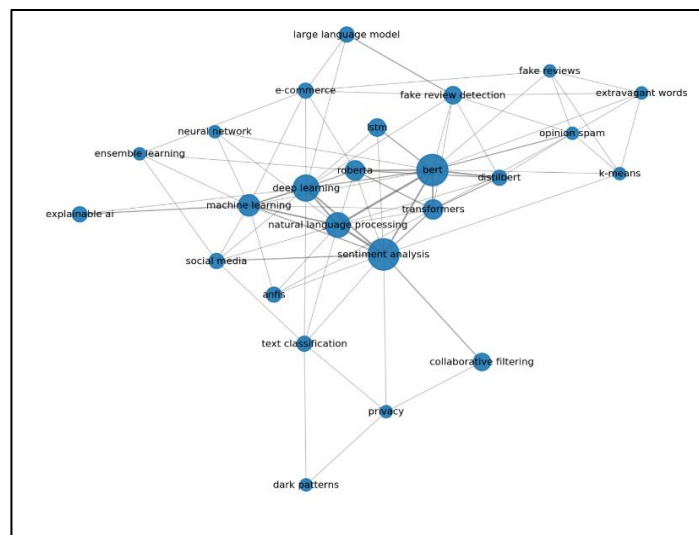


Figure 1: Keyword co-occurrence network (top 25 author keywords) derived from 76 Scopus-indexed studies.
Source: Authors, (2026).

II.1 DECEPTIVE AND FAKE REVIEW DETECTION

Early studies present opinion spam as a threat to online commercial trustworthiness, prompting the development of linguistic, behavioral, and metadata-based classification approaches [2-4]. Thereafter, deep learning architectures have replaced manually selected features with significant advances in transformer-based deception detection [5-7]. Ennaouri and Zellou proposed an influential model leveraging linguistic exaggeration cues, BERT embeddings, and fuzzy logic, demonstrating performance improvements [5]. A few researchers proposed and tested algorithms based on ensemble multidimensional feature selection to improve fake review identification [8], [9].

II.2 SENTIMENT ANALYSIS AND RATING PREDICTION

Early research in sentiment analysis showed that the emotions people express in their reviews helps in accurate predictions of rating. When a recommender system accurately understands whether a user feels positive or negative, it can better guess product ratings and user preferences [10]. Newer models, like transformers, use sentiment dictionaries to identify sentiments. Transformers worked across multiple languages to achieve even better results [11], [12]. Recent work on e-commerce platforms shows that after adding lexicon-based features to BERT-style models, improved the accuracy in detection of sentiment [13]. Other research focused on how large language models can analyze sentiment in more than one language on different digital platforms [11].

II.3 DARK PATTERNS AND DECEPTIVE INTERFACE DESIGN

The literature not only also discusses research focused on misleading text but also on “dark patterns” in user interfaces. Dark patterns are design tricks that push users to act in certain ways by hiding information or using visually manipulative layouts. Studies show that many cookie consent pop-ups use deceptive designs that make it hard for users to understand their choices or maintain control [14]. Other researchers are developing ways to automatically detect these deceptive UI practices through methods such as explainable AI (XAI), logs of how users interact with websites, and pattern mining techniques to identify patterns used in deceptive UI [15], [16].

II.4 CHAOS THEORY, COMPLEX SYSTEMS, AND ANOMALY DETECTION

Chaos theory and non-linear dynamics have been used by researchers in information systems to find non-linear and unpredictable adaptive behaviors on digital platforms [17-19]. Methods like higher-order cumulants, phase-space reconstruction, and Lyapunov exponent analysis help in identifying hidden patterns and irregular information flows in the system [19], [20]. More recent applications of chaos theory include finding anomalies in cloud systems through local explainability and multiscale analytics [21]. To show what the studies focus on and how the research has changed over time, Figure 2 presents the relationship between evidence screening and the development of themes in the research. The combined PRISMA–Topic Evolution visualization highlights not only the full set of 76 studies, but also how research priorities have shifted in response to new challenges in digital information environments. The figure illustrates the evidence screening process followed by PRISMA 2020. The process identifies 76 Scopus records. After screening these 76 records, removing 0, assessing 50 full texts meta-analysis revealed that the topic evolution shows a move from complexity and chaos modeling (1991–2005) to textual fraud and sentiment extraction (2010–2018), and then toward LLM-based fake review detection and XAI-supported dark pattern auditing (2023–2026). It also shows hybrid crossover themes such as chaos–anomaly synergy, sentiment–deception fusion, LLM–sentiment trust systems, and XAI–manipulative design auditing.

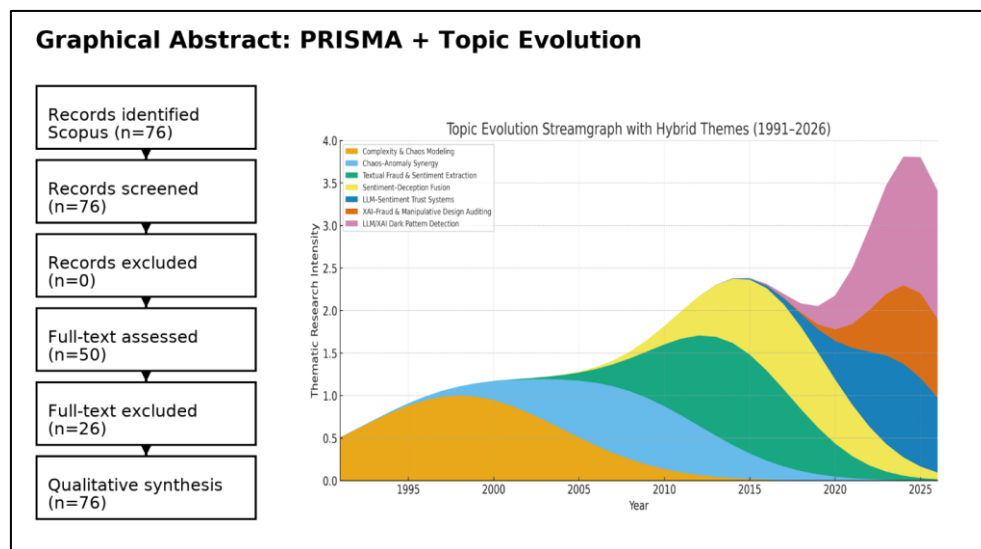


Figure 2: Combined PRISMA–Topic Evolution Graphical Abstract.
Source: Authors, (2026).

II.5 CROSS-CUTTING RESEARCH TRENDS

The review of all 76 studies shows that there are five main research trends that appear across different topics and methods. These trends show how the research has evolved over time. These five directions found across the literature are summarized below:

- Transformer/LLM-driven analytics [5-7], [13], [22], [11].
- Hybrid feature engineering remains essential [13], [10], [11], [23].
- XAI and algorithmic transparency gain prominence [14-16], [21].
- Multilingual modeling and cross-domain transfer challenges persist [13], [7], [11], [12], [24].
- Platform governance and trust [17], [18], [25].

The literature survey shows that research in this area is evolving fast and is driven by tools for analyzing text, detecting manipulative interfaces, and modeling complex systems. For example, transformer models have improved classification accuracy, but more research is required to fill the important gaps especially in explainability, cross-platform interoperability, and connecting the research to public policy.

III. METHODOLOGY

In this section, we explain the methodology followed in calculating the NETER score. The steps include estimating sentiments, detecting deception, applying confidence weights, measuring and assessing instability through chaos-theory. By following these steps raw e-commerce reviews are converted into a strong, manipulation-resistant trust score for a product. The process begins with collecting and preprocessing data.

III.1 DATA COLLECTION AND PREPROCESSING

Here, we explain how the data for the study is collected, organized, and prepared. To make sure that the data is accurate, consistent, and relevant we followed a clear and structured process at every step of research.

III.1.1 Data Source

In this study, we created a dataset for product reviews written by users on Amazon e-commerce platform. The data consisted of 157 text reviews of a product. Each review is a short piece of text written by a customer. Each record in the dataset has textual review, time at which the review is posted along with the numerical value of star rating given by the customer. The dataset reflects real customer behavior as review text may have excessively hyped language, possible misleading design cues, and changes in sentiments. These features are important to test the working of the NETER framework.

III.1.2 Data Cleaning

In this step the reviews are cleaned to make sure that the data is consistent for applying chaos-based modeling and reliable for evaluation of sentiments, deception pattern checks, and stability measurement. Any empty or non-text entries are removed so that the dataset remains accurate and useful for evaluation.

III.2 ROBERTA-BASED SENTIMENT ANALYSIS

The first step in evaluation of sentiments is to perform sentiment analysis where text is analyzed by using the transformer-based model, RoBERTa. This model has been trained to identify and classify sentiments. RoBERTa divides each review into tokens and gives raw scores (logits) for three classes: negative, neutral, and positive. The SoftMax layer of the model converts these scores into probabilities. The model outputs logits for classes $[neg, neu, pos]$. Let the probability associated with the positive class be denoted as $P_{positive}$. The SoftMax () is applied on pos class to obtain $P_{positive}$ as shown in Equation 1.

$$P_{positive} = SoftMax(pos) \quad (1)$$

NETER focuses on positive sentiments because they are most manipulated in e-commerce reviews. Positive reviews influence buying decisions, and dark patterns typically inflate positivity rather than negativity. Since negative reviews are seldom faked, NETER evaluates how trustworthy positive sentiment is, rather than treating all sentiments equally. $P_{positive}$ is directly used in the NETER computation, replacing prior polarity-based approximation methods and providing more context-aware sentiment estimation as compared to earlier, polarity-based estimates, which relied on simple methods that classified text as positive, negative, or neutral using word lists or basic rules. The traditional approaches counted positive or negative words without understanding context, sarcasm, emphasis, or nuanced phrasing. As a result, they often produced oversimplified and inaccurate sentiment scores.

III.3 DECEPTION AND DARK-PATTERN DETECTION

To estimate how likely a review is using manipulative or exaggerated language, we use a deception-detection module based on linguistic cues. The text in each review is checked for usage of superlatives (e.g., “amazing” or “perfect”), promotional phrases (e.g., “#1 bestseller,” “buy now”), too much punctuation, ALL-CAPS words, very short but extremely positive comments, and urgency-based words (e.g., “limited period offer”). These signals are combined into a weighted score, which is then passed through a logistic function to calculate the dark-pattern probability using Equation 2.

$$P_{dark} = \sigma(z - 0.6) \quad (2)$$

The Equation 3 is used to calculate total deception score for a product by averaging dark-pattern probabilities across all reviews.

$$P'_{dark} = \frac{1}{n} \sum_{i=1}^n P_{dark_i} \quad (3)$$

III.4 REVIEW CONFIDENCE WEIGHT

To estimate the reliability of sentiments based on the number of available reviews, a diminishing return weighting function is applied. The confidence weight of reviews is defined using Equation 4.

$$w(n) = 1 - e^{-n/k} \quad (4)$$

The n in Equation 4 represents the number of reviews for a product and k is a scaling constant (typically set to 50). This method considers that products with only a few reviews are given lower confidence weight in comparison to products with more reviews which are assigned higher weights. This prevents small amounts of data from having too much impact on the trust evaluation.

III.5 CHAOS-THEORY ADJUSTMENT (LYAPUNOV EXPONENT)

To measure nonlinear instability or increased hype in review patterns, the product’s sentiment-deception score is used as the starting point in a logistic map, and this value is calculated using Equation 5:

$$x_0 = P'_{positive}(1 - P'_{dark}) \quad (5)$$

The value of chaos-control (r) parameter is initialized using Equation 6 and is used to iterate the logistic map.

$$r = 3.2 + 0.8 \cdot P'_{dark} \quad (6)$$

The chaos-control parameter allows products with higher deception probabilities to shift toward the chaotic regime. The value of r indicates that products with higher average dark-pattern probability are closer to the chaotic region of the logistic map. In other words, r controls how sensitive the logistic map is to changes in a product's sentiment patterns and is illustrated using Figure 3. The Figure 3 shows that as deception or hype increases in product reviews, the logistic-map parameter r shifts towards right, representing a transition from stable and authentic review behavior to unstable and manipulated sentiment patterns.

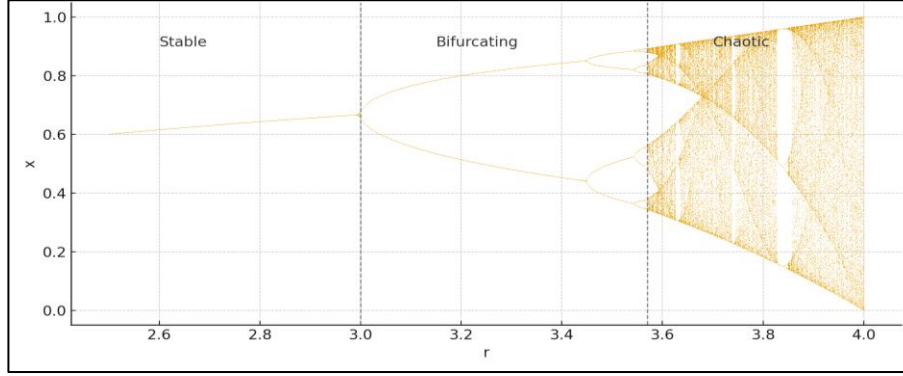


Figure 3: Simplified Logistic Map Bifurcation (NETER Zones).

Source: Authors, (2026).

In the stable zone ($r < 3.0$), reviews are consistent and predictable. For example, a kitchen blender review might receive balanced feedback such as: “Works well for everyday use” or “Good value for the price” in majority of the cases. In this zone there are minor variations in the sentiment patterns, but in general sentiment remains consistent and reflects real user experience. In the bifurcation zone ($3.0 \leq r < 3.57$), sentiment patterns begin to oscillate. This indicates fluctuations in the sentiments of the users, divided responses, or marketing influence e.g. one review of smartwatch may be an extremely positive comment after a promotional campaign such as “Absolutely life-changing! Must buy!” followed by negative reviews such as “Battery died in two days, not worth it.” This irregularity in reviews indicates nonlinear shifts in trust and is similar to repeating patterns appearing in the logistic map. The system enters the chaotic zone when value of $r = 3.57$. This indicates sharp fluctuations in emotional tones of sentiment. For example, an influencer advertising a beauty product may get a mix of contradictory emotions in reviews where a few reviews may be extremely positive while a few may be extremely negative. These reviews appear back-to-back and change very quickly. For example, “Miracle results in one night!!!” or “Total scam, it burned my skin”. This type of instability represents possible manipulation. NETER uses chaos theory to detect this instability and awards a penalty to such reviews by reducing the trust score. The value of $r = 3.2$ represents P'_{dark} is low and predictable system behavior [26]. The logistic map is iterated using the Equation 7 and the Lyapunov exponent is computed using Equation 8.

$$x_{t+1} = rx_t(1 - x_t) \quad (7)$$

$$\lambda = \frac{1}{T} \sum_{t=1}^T \ln |r(1 - 2x_t)| \quad (8)$$

The Lyapunov exponent shows fluctuation in the product sentiments. If small changes in reviews quickly lead to big shifts in overall sentiment, the value of the exponent will be high. A high value means the emotional tone in reviews is unstable, while a low value means the emotional sentiments are consistent and predictable. The chaos penalty is normalized by using Equation 9. The $clip()$ in the equation normalizes value of $\lambda / (\ln 2)$ within $[0,1]$.

$$ChaosPenalty = clip(\lambda / (\ln 2), 0, 1) \quad (9)$$

III.6 REVIEW STABILITY MEASUREMENT

The consistency and reliability of the reviews is checked in two steps. First, the sentiment scores from the reviews are grouped into five equal ranges, from very negative to very positive. Then, Shannon entropy, $H = -\sum p_i \ln p_i$, is calculated, which is normalized using formula $H_n = H / \ln(5)$, to check the similarity and differences in the sentiment patterns. If reviews fall evenly across the five ranges, the entropy is high, meaning the opinions are mixed whereas reviews in one or two ranges show low entropy and consistency in the emotional tone of reviews. Further the model uses Median Absolute Deviation (MAD) to measure the spread of review sentiments. This spread tells us whether reviews are similar or not. The value of MAD is normalized to $MAD_n = MAD / (0.25 / 0.1)$ so that the spread fits between values 0 and 1. A low value means the reviews are close together, while a high value means they are widely scattered. This measure, MAD_n , along with the normalized entropy H_n , is combined to form the overall Stability Factor, denoted as *Stability*, and is calculated using Equation 10.

$$Stability = clip((1 - 0.5H_n)(1 - 0.5 MAD_n), 0.5, 1.0) \quad (10)$$

The above equation ensures that the final value is between 0.5 and 1.0. The Stability Factor ensures that normal differences in user opinions do not lower the trust score because in the real-world people naturally think differently. *Stability* values below 0.5 will penalize entropy, while values above 1.0 would inflate trust. This range between 0.5 and 1.0 ensures that the stability factor moderates, rather than dominates, the overall NETER rating. Products with higher Stability values show consistent and reliable sentiments in textual reviews.

III.7 NETER SCORE FORMULATION

The final trust score or NETER score integrates all components and is calculated using Equation 11:

$$NETER = 100 \cdot w(n) \cdot P_{positive} \cdot (1 - P_{dark}) \cdot (1 - ChaosPenalty) \cdot Stability \quad (11)$$

The final trust score calculated by NETER lies between 0 and 100. A higher NETER score means the product's reviews are more reliable, consistent, and less likely to be influenced by hype, manipulation, or deceptive patterns. A low NETER score alerts customers that the reviews may be unreliable or inconsistent and may contain deceptive patterns. Figure 4 summarizes methodology for calculating the NETER score.

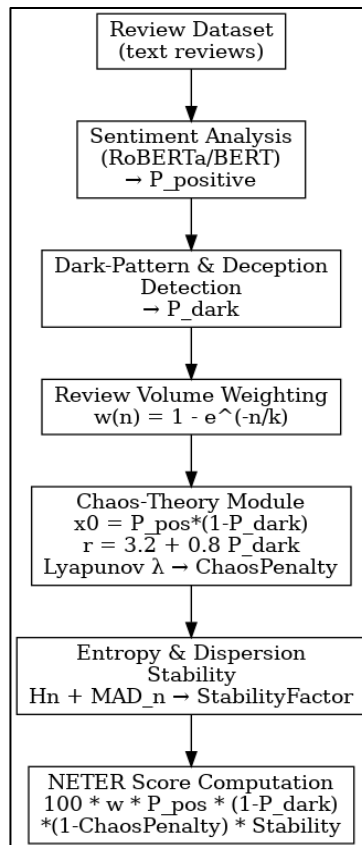


Figure 4: Steps to compute NETER score.

Source: Authors, (2026).

IV. RESULTS AND DISCUSSIONS

In this study a dataset of 157 Amazon product reviews is created. The NETER model is applied on this dataset for understanding the reliability of customers' reviews. The model identifies the emotions expressed in reviews by using a LLM model, RoBERTa and checks for manipulative or promotional language. It also finds chaotic shifts in review patterns and calculates stability or consistency in opinions of users over time. The NETER model begins with calculation of $P_{positive}$, P_{dark} and confidence weights and is illustrated in the next section.

IV.1 EVALUATION OF $P_{positive}$, P_{dark} AND CONFIDENCE WEIGHT

The review dataset shows that the average value of $P_{positive}$ is 0.6398 which indicates that reviews of customers are generally positive. The histogram shown in figure 5 indicates the distribution of $P_{positive}$ of all reviews in the dataset. This histogram is right skewed and widely spread. The values of $P_{positive}$ ranges between 0.4 and 0.8, indicating many reviewers have a positive view of the product. The highest concentration of $P_{positive}$ is between 0.55 and 0.70. A spike closer to 1.0 indicates a large number of positive reviews from enthusiastic customers or promotional language. A few reviews fall in the low range of 0 – 0.2, showing that some customers have negative experiences. This wide spread of values results in a high entropy score ($H_n = 0.813$), which means sentiments are inconsistent. Instead, they sharply vary, suggesting that user experiences are inconsistent or unstable.

The language-cue analysis also found many reviews with persuasive or exaggerated emotional tone are often related with dark-pattern behavior. An average deceptive-language probability of $P_{\square\square\square\square} = 0.4762$ means nearly half the reviews show some signs of potentially misleading phrasing and a high review confidence weight of $w(n) = 0.9567$, shows that the sentiment, deception, and stability measures are based on a sample large enough to be dependable.

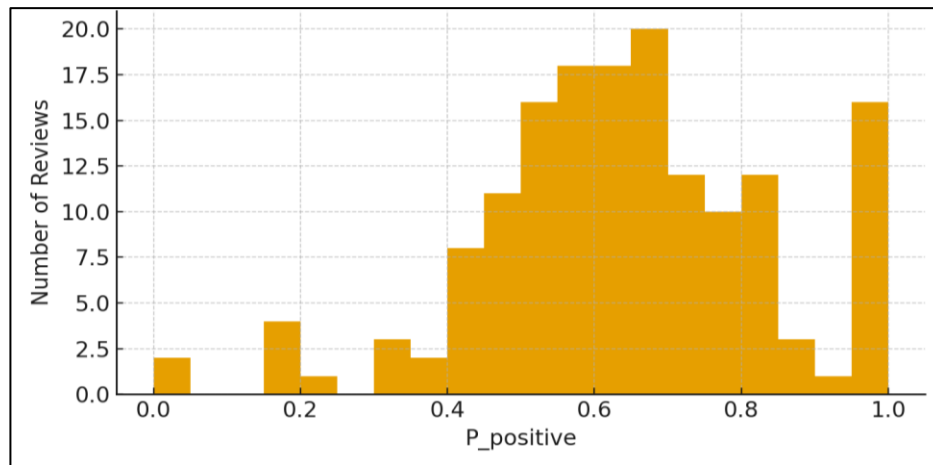


Figure 5: Sentiment Distribution (P_positive).
Source: Authors, (2026).

IV.2 CHAOS AND STABILITY ASSESSMENT

To measure how sentiment patterns might shift towards chaotic regions, the logistic-map-based Lyapunov exponent is calculated, giving a value of $\lambda = 0.1082$. This value corresponds to a *ChaosPenalty* of 0.1561. In other words, the reviews show signs of temporary hype spikes and back-and-forth fluctuations, but they do not behave in a fully chaotic or random manner. Figure 6 shows the Chaos Dynamics plot, where the logistic-map value x_t iterates over 120 iterations using parameters taken from the review data. With a chaos-control parameter of $r = 3.581$, the system shifts into a weakly chaotic state. This is supported by the positive Lyapunov exponent ($\lambda = 0.1082$), which indicates that small changes in sentiment can grow over time. The resulting *ChaosPenalty* of 0.1561 reflects the moderate but noticeable level of sentiments instability.

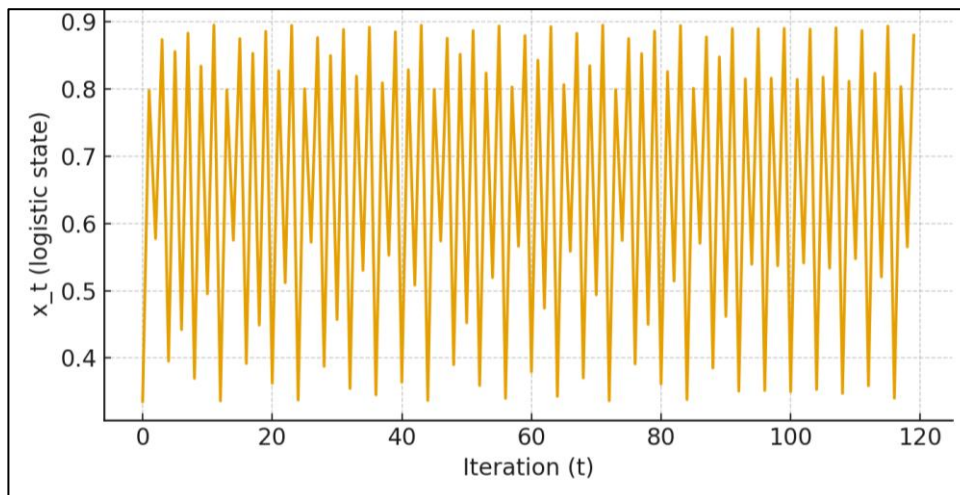


Figure 6: Chaos Dynamics.
Source: Authors, (2026).

IV.3 ASSESSMENT OF ENTROPY, DISPERSION AND STABILITY

The entropy and dispersion methods are used to study distribution of sentiment patterns in the review dataset. NETER model calculates the normalized Shannon entropy ($H_n = 0.8126$) by dividing the predicted values of sentiment into equal-sized intervals. This metric indicates uniformity of the distribution of reviews over different sentiment patterns. A higher value of entropy indicates variation in emotional tone of reviews, where reviewers have a mix of positive and negative experiences. Next, the normalized Median Absolute Deviation ($MAD_n = 0.5278$) is calculated to measure the difference in individual review sentiments from the average $P_{\square\square\square\square\square\square}$ value. A higher MAD value indicates disagreement among reviewers about their experience of using a product. Both measures are then combined in the NETER stability formula where the stability score is reduced when the sentiment distribution shows a lot of variation or inconsistency. After applying this penalty, the Stability Factor is calculated as 0.5000. This value of Stability Factor indicates that reviews do not follow a steady or unified pattern instead, they show variation in emotional tone of reviews. Even though the sentiment is positive, the instability reduces the trust rating for the product because the product's feedback is inconsistent, unpredictable, and less dependable.

IV.4 NETER TRUST SCORE

In this study, we measured the reliability of a product's reviews by calculating the trust score. The average star rating is also considered along with trust score. In our study the star rating of the product is around 4 out of 5, which indicates high trust, but the final NETER trust score is only 13.53 out of 100 which is very low. Even though the overall sentiment and star ratings are positive, a low NETER score indicates that reviews may be untrustworthy.

IV.5 INTERPRETATION OF FINDINGS

The results obtained from the proposed model indicate that positive ratings can be misleading, unstable review patterns reduce credibility, and high variation in the reviews makes reviews of the product less trustworthy. These points are summarized below.

1. **Positive words in reviews do not always indicate trustworthy feedback:** Nearly half of the reviews showed signs of promotional manipulation, and sentiment fluctuation between reviews.
2. **Chaos and instability are important factors of review credibility:** When reviews quickly change from positive to negative, they show polarized or coordinated reviews.
3. **Entropy-based stability is a strong penalizing factor:** High entropy and dispersion resulted in low Stability Factor and hence, low trust score.

Our findings show that the NETER model gives a better measure of review credibility than methods that only look at sentiments. It also helps in revealing the presence of manipulation. The NETER score shows the difference between the actual and the perceived trust of users. As per our results, the average sentiment score ($P_{\text{sentiment}} = 0.6398$) shows that large number of users have positive experience with the product but value of $P_{\text{manipulation}} = 0.4762$ indicates that manipulative language or deceptive patterns are present in reviews. $P_{\text{manipulation}}$ considers manipulations like urgency wording, exaggerated praise, or repeated hype which are part of the company's marketing tricks. The entropy ($H_n = 0.8126$) and dispersion ($MAD_n = 0.5278$) values also indicate inconsistency in the user's reviews where a few may be artificially generated reviews to influence buyers. This study finds that the quick change in positive and negative sentiments may appear before promotional activities. The Lyapunov exponent ($\lambda = 0.1082$) and derived *ChaosPenalty* shows that the review dataset has nonlinear instability. This supports the idea that the chaos-based model can be used for detecting irregular patterns and artificial boosts on e-commerce platforms. Traditional sentiment-only models would likely classify this product as positively perceived (with 4-star rating), yet NETER calculates a low final trust score of 13.53/100, demonstrating that positivity is not a sufficient indicator of authenticity. These findings reinforce the claim that credibility assessment requires a multidimensional analysis that accounts for sentiment, manipulation, and dynamical behavior.

V. CONCLUSIONS

NETER combines transformer-based sentiment analysis, dark-patterns and deception detection, nonlinearity detection using chaos theory, and entropy-based stability scoring to evaluate credibility of the reviews. It produces more accurate and manipulation-resistant trust metrics compared to scores which are based only on sentiment or product-rating. This analysis shows that traditional sentiment and star ratings fail to capture deceptive language, instability and fake sentiment amplification in reviews. Highly positive reviews alone cannot indicate trust when entropy and chaos are elevated. NETER overcomes above-mentioned limitations by combining sentiment analysis, identification of deceptive patterns, stability, and nonlinear dynamic modelling to provide a score which is more reliable and trustworthy. When we applied NETER on our dataset, we obtained low trust score of 13.53/100, despite positive sentiments in reviews. This confirms that positive sentiments in reviews can mask manipulation which includes deceptive promotional strategies that can influence consumers. A low trust score of 13.53 means that reviews for that product are not reliable and trustworthy. It also means that the comments are unstable, exaggerated, manipulated, and the overall sentiment is inconsistent. Thus, NETER identifies deceptive or exaggerated language in reviews, fake reviews and unreliable reviews by highlighting nonlinearity and provides better understanding of how trust is influenced by paid promotions, sponsored products or fake reviews.

In future work, we are planning to compare reviews of a product available on different ecommerce websites and apps that can help in identifying more reliable shopping platforms for the customers. Moreover, NETER can also be embedded in real time within e-commerce platforms to monitor and prevent fraud using fake reviews. This will help in detection and mitigation of deceptive patterns in product reviews, generating more reliable trust scores.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Preeti Marwaha, Arunita Chaukiyal.

Methodology: Preeti Marwaha, Arunita Chaukiyal.

Investigation: Preeti Marwaha, Arunita Chaukiyal.

Discussion of results: Preeti Marwaha, Arunita Chaukiyal.

Writing – Original Draft: Preeti Marwaha, Arunita Chaukiyal.

Writing – Review and Editing: Preeti Marwaha, Arunita Chaukiyal.

Resources: Preeti Marwaha, Arunita Chaukiyal.

Supervision: Preeti Marwaha, Arunita Chaukiyal.

Approval of the final text: Preeti Marwaha, Arunita Chaukiyal.

VII. REFERENCES

- [1] N. J. van Eck and L. Waltman, "Visualizing bibliometric networks," in *Measuring Scholarly Impact: Methods and Practice*, Cham, Switzerland: Springer, 2014, pp. 285–320.
- [2] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," *arXiv preprint*, arXiv:1107.4557, 2011.
- [3] S. Feng, L. Xing, A. Gogar, and Y. Choi, "Distributional footprints of deceptive product reviews," in *Proc. Int. AAAI Conf. Web and Social Media (ICWSM)*, vol. 6, no. 1, 2012, pp. 98–105.
- [4] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *J. Big Data*, vol. 2, no. 1, p. 23, 2015.
- [5] M. Ennaouri and A. Zellou, "Enhancing fake review detection using linguistic exaggeration, BERT embeddings, and fuzzy logic," *IEEE Access*, 2025.
- [6] A. Filali and M. Merras, "Enhancing spam detection with GANs and BERT embeddings: A novel approach to imbalanced datasets," *Procedia Comput. Sci.*, vol. 236, pp. 420–427, 2024.
- [7] D. A. Bumber, F. Z. Qachfar, and R. Verma, "Domain-agnostic adapter architecture for deception detection: Extensive evaluations with the DIFrauD benchmark," in *Proc. Joint Int. Conf. Comput. Linguistics, Lang. Resources and Evaluation (LREC-COLING)*, 2024, pp. 5260–5274.
- [8] S. Li *et al.*, "A deceptive reviews detection method based on multidimensional feature construction and ensemble feature selection," *IEEE Trans. Comput. Social Syst.*, vol. 10, no. 1, pp. 153–165, 2022.
- [9] Y. W. Oh and C. H. Park, "Machine cleaning of online opinion spam: Developing a machine-learning algorithm for detecting deceptive comments," *Amer. Behav. Sci.*, vol. 65, no. 2, pp. 389–403, 2021.
- [10] D. Yu, Y. Mu, and Y. Jin, "Rating prediction using review texts with underlying sentiments," *Inf. Process. Lett.*, vol. 117, pp. 10–18, 2017.
- [11] M. R. A. Rashid *et al.*, "BERT-KAN: Enhancing bilingual sentiment analysis in Bangladeshi e-commerce through fine-tuned large language models," *Natural Language Processing Journal*, p. 100190, 2025.
- [12] S. Rani and P. Kumar, "Aspect-based sentiment analysis using dependency parsing," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 21, no. 3, pp. 1–19, 2021.
- [13] L. P. Manik, H. Susianto, A. Dinakaramani, N. Pramanik, and T. Suhardijanto, "Can lexicon-based sentiment analysis boost performances of transformer-based models?" in *Proc. 7th Int. Conf. New Media Studies (CONMEDIA)*, 2023, pp. 314–319.
- [14] E. Kocyigit, A. Rossi, and G. Lenzini, "Towards assessing features of dark patterns in cookie consent processes," in *Proc. IFIP Int. Summer School Privacy and Identity Management*, Cham, Switzerland: Springer, 2022, pp. 165–183.
- [15] U. Ehsan and M. O. Riedl, "Explainability pitfalls: Beyond dark patterns in explainable AI," *Patterns*, vol. 5, no. 6, 2024.
- [16] P. Hájek and J.-M. Sahut, "Mining behavioural and sentiment-dependent linguistic patterns from restaurant reviews for fake review detection," 2022.
- [17] N. McBride, "Chaos theory as a model for interpreting information systems in organizations," *Inf. Syst. J.*, vol. 15, no. 3, pp. 233–254, 2005.
- [18] M. Harle and U. Feudel, "On the relation between predictability and homoclinic tangencies," *Int. J. Bifurcation Chaos*, vol. 15, no. 8, pp. 2523–2544, 2005.
- [19] S. H. Lang *et al.*, "A study on methods for determining phase space reconstruction parameters," *J. Comput. Nonlinear Dyn.*, vol. 17, no. 1, p. 011006, 2022.
- [20] C. Wang *et al.*, "Analysis of the time series generated by a new high-dimensional discrete chaotic system," *Complexity*, vol. 2018, Art. no. 9818520, 2018.
- [21] W. Villegas-Ch *et al.*, "Integrating explainable artificial intelligence in anomaly detection for threat management in e-commerce platforms," *IEEE Access*, 2025.
- [22] M. Leon, "Sentiment analysis: From rule-based lexicons to large language models," *Intelligent Systems with Applications*, p. 200599, 2025.
- [23] S. Xu *et al.*, "A hybridized approach for enhanced fake review detection," *IEEE Trans. Comput. Social Syst.*, vol. 11, no. 6, pp. 7448–7466, 2024.
- [24] A. S. Ipa *et al.*, "BDSentiLLM: A novel LLM approach to sentiment analysis of product reviews," *IEEE Access*, 2024.
- [25] B. Liu, *Sentiment Analysis and Opinion Mining*. Cham, Switzerland: Springer Nature, 2022.
- [26] M. Rani and R. Agarwal, "A new experimental approach to study the stability of logistic map," *Chaos, Solitons & Fractals*, vol. 41, no. 4, pp. 2062–2066, 2009.