Real Time Detection of Evolving Spammer Groups in E-Commerce

Talluri Ramesh Babu[1], Madineni Nandini[2], Talasila Greeshma[3], Vajrala Narendra Reddy[4]
[1,2,3,4]*Vasireddy Venkatadri Institute of Technology*

***Abstract*—E-commerce platforms are increasingly affected by spammer groups that manipulate product reviews to influence ratings and mislead customers. Detecting such coordinated activities is challenging, especially when spammers continuously change their behaviour. In this work, we propose a hybrid detection framework that combines graph-based modelling, behavioural analysis, and machine learning techniques to identify spammer groups. The system constructs a user product interaction graph to capture relationships among reviewers and applies TF-IDF with cosine similarity to analyse textual patterns. Community detection is used to identify coordinated reviewer groups, while a Random Forest model is used for classification. The results show that the proposed approach can effectively detect suspicious users and groups with improved accuracy and reduced false positives. This makes the system suitable for practical use in large-scale e-commerce platforms.**

***Index Terms*—Spammer Group Detection, E-commerce Reviews, Graph-based Modelling, Behavioural Feature Analysis, Machine Learning Classification, Real-time Spam Detection, Community Detection**

I. INTRODUCTION

1.1 Rise of Spammer Group Activities in E-Commerce
Online reviews significantly influence customer trust and product reputation, but spammer groups exploit this by posting deceptive and coordinated reviews to manipulate ratings and mislead consumers [1], [6]. Traditional detection methods such as clustering and rule-based analysis mainly focus on individual reviewers and struggle to identify evolving, coordinated group-level attacks [3], [4], [10]. In real-world scenarios, spammer groups often coordinate their activities in subtle ways, making them difficult to detect using traditional methods that focus only on individual users. To overcome these challenges, we propose a hybrid detection framework that combines graph-based modeling, behavioural analytics, and machine learning techniques [7], [8]. The system model's user product interactions as graphs to uncover hidden collusive patterns, while behavioural and textual features such as review frequency, rating behaviour, temporal bursts, and content similarity are analysed.[1], [9]. Furthermore, existing techniques face limitations including high computational cost, poor scalability, and reduced accuracy when applied to large e-commerce platforms [2], [5]. In this framework, user reviews are aggregated into reviewer product interaction data and modelled as dynamic graphs to capture relational patterns, enabling effective detection of coordinated spammer groups. [3], [7], [10].

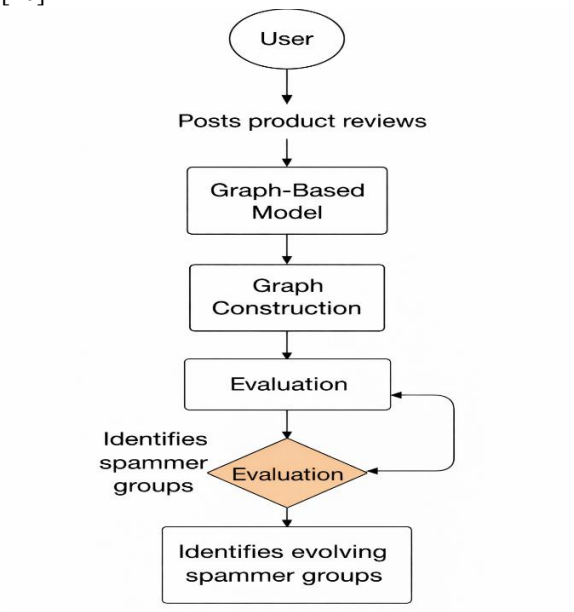

Fig 1. Workflow of Spammer Group Detection

Figure 1 illustrates the overall workflow of spammer group identification in the proposed system. The process begins with users posting product reviews on

the e-commerce platform, which are collected as input data. These reviews are processed using a graph-based model that represents user–product interactions to capture hidden relational and behavioural patterns. The constructed graph is then evaluated to identify coordinated activities among reviewers. Through iterative evaluation and feedback, the system effectively detects spammer groups and tracks their evolving behaviour over time. This workflow highlights how graph construction and evaluation play a crucial role in identifying both existing and evolving spammer groups.

1.2. Limitations of Existing Systems

Despite significant progress in spammer detection research, several limitations persist in existing systems. Most current approaches heavily rely on labeled datasets, making them less adaptable to emerging spammer behaviours and reducing their generalization capability [1], [6]. Scalability also remains a challenge, as graph-based models often produce highly dense structures that lead to computational overhead when applied to large-scale e-commerce platforms [2], [7]. Moreover, spammer groups continuously evolve their strategies—altering posting patterns, language styles, and coordination techniques—causing static models to degrade over time due to concept drift [3], [5]. The presence of noisy and ambiguous data further complicates detection, as legitimate coordinated user activities can often resemble spammer behaviours, resulting in high false positives [4], [8].

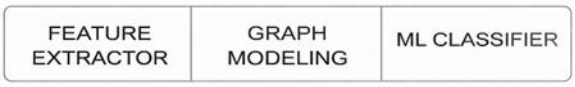

Fig 2. System Architecture

Additionally, advanced deep and graph-based learning methods, though powerful, suffer from limited interpretability, making it difficult for analysts to understand or justify model predictions [9], [10]. Finally, practical deployment issues such as data privacy restrictions, limited access to real-time behavioural logs, and challenges in maintaining low-latency detection hinder the effective real-world adoption of existing spammer group detection systems[6],[12].

1.3. Research Contributions

The primary contributions of this research are threefold. First, we propose a novel hybrid detection framework that combines graph-based modelling, behavioural analysis, and machine learning to identify both individual and group-level spammer activities in real time [1], [3], [6]. Second, we introduce a comprehensive feature extraction pipeline that integrates behavioural, textual, and temporal attributes, effectively capturing coordinated spam patterns such as rating bursts, review similarity, and posting synchronization [2], [4], [7]. Third, the system is designed in a modular way so that the detection model can be updated and retrained when new review data becomes available [5], [9], [1] Additionally, the system provides explainable insights through interpretable model outputs and automated reporting mechanisms, enabling transparency and aiding e-commerce administrators in decision-making [6], [8]. This end-to-end framework enhances detection accuracy, reduces false positives, and bridges the gap between research prototypes and real-world deployment in large-scale online platforms [7], [11], [12].

1.4. Motivation of the Proposed Framework

Spammer groups in e-commerce are becoming increasingly sophisticated, using coordinated behaviours, fake identities, and adaptive review patterns to manipulate product reputations and mislead customers [1], [3], [6]. Existing detection systems struggle to keep up due to limited scalability, static learning models, and poor adaptability to evolving spammer strategies [2], [5], [7]. This motivates the development of a real-time, intelligent, and explainable detection framework that integrates graph-based analysis, behavioural modelling, and machine learning techniques [6], [8], [9]. The proposed system aims to enhance detection accuracy, reduce false positives, and maintain transparency in decision-making, ensuring a trustworthy and scalable solution for modern e-commerce environments [10], [11], [12].

II. LITERATURE REVIEW

Spammer group detection in e-commerce has gained increasing attention due to the rising impact of fake and coordinated reviews on consumer trust. Traditional techniques such as clustering, rule-based

analysis, and temporal filtering often fail to detect evolving and large-scale spammer behaviours. Recent studies highlight the effectiveness of graph-based modeling and machine learning algorithms in capturing reviewer relationships and identifying collusive groups. Integrating behavioural, textual, and temporal features has been shown to significantly improve detection accuracy and scalability, paving the way for more adaptive and intelligent detection frameworks.

Yang et al. [1] further showed that group-level similarity analysis is more effective than individual reviewer profiling for detecting coordinated spam. However, their method is sensitive to minor content variations and struggles with detecting evolving spammer strategies. Their framework mainly focuses on textual duplication and overlooks deeper behavioural coordination signals. As a result, it becomes less effective against spammers who use paraphrased or AI-generated reviews. Dhawan et al. [2] demonstrated that Reviewer2Vec embeddings improve collective fraud representation in graphs. Nevertheless, the approach requires significant computational resources and does not scale efficiently for real-time large datasets. The model performs well in capturing latent reviewer relationships through embeddings. However, it requires frequent retraining to remain effective against newly emerging spammer strategies. Wang et al. [3] highlighted the effectiveness of Markov Random Fields in modelling relational dependencies among reviewers. However, their probabilistic framework involves high inference complexity and lacks adaptability to dynamic spammer behaviours. Their use of probabilistic inference improves relational reasoning across users. Yet, the approach is sensitive to noise in temporal data and may suffer from delayed detection.

Li et al. [4] proved that co-bursting and temporal behaviour patterns are strong indicators of coordinated spam. Their approach, however, depends on fixed time windows and performs poorly when spammers change posting intervals. The co-bursting model effectively identifies short-term coordinated attacks. However, it struggles to capture long-term collusive behaviours spread over extended time periods.while Wang et al. [5] introduced reinforcement learning to detect overlapping spammer groups more accurately. Despite its effectiveness, the model is computationally expensive and difficult to deploy in real-time systems. The reinforcement learning agent improves adaptability to dynamic spammer patterns. Still, the training process is time-consuming and requires careful hyperparameter tuning.

Mukherjee et al. [6] showed that graph-based ranking can successfully identify early-stage spammer communities. However, the approach primarily relies on structural patterns and ignores semantic similarities in review content. The ranking-based approach enables early warning of potential group spam. However, it does not incorporate temporal dynamics, limiting its effectiveness for evolving attacks. while Akoglu et al. [7] demonstrated that dense subgraph detection is highly scalable for large review networks. Still, the method may misclassify legitimate dense communities as spammer groups, leading to higher false positives. The algorithm efficiently detects dense suspicious blocks in large graphs. Nevertheless, it lacks contextual awareness and may flag legitimate promotional campaigns as spam. Jiang et al. [8] found that frequent itemset mining effectively reveals overlapping co-review behaviours. However, its performance degrades on large-scale datasets and does not incorporate temporal evolution modeling. The method performs well for identifying tightly knit reviewer clusters. However, it fails to generalize across domains with highly diverse product categories. Hooi et al. [9] improved anomaly detection accuracy by enhancing dense block discovery in bipartite graphs. Yet, the model lacks explainability and struggles with adaptive learning for evolving spam patterns. The enhanced detection mechanism improves robustness against noisy data. Still, it does not support online learning for continuously arriving review streams. Xie et al. [10] showed that temporal synchronization analysis enables early detection of coordinated review bursts. Nevertheless, their approach does not integrate textual semantics and is limited in handling long-term behaviour drift. The early-burst detection strategy helps prevent large-scale review manipulation. However, it does not consider cross-product coordination, reducing group-level detection accuracy.

Liu et al. [12] propose a hybrid framework that combines large language model (LLM) embeddings with graph neural networks to detect AI-generated spam reviews. Their approach captures both semantic patterns in review text and relational structures among reviewers, enabling improved detection accuracy. The

study demonstrates strong performance against modern AI-generated reviews that bypass traditional detectors. However, the model requires substantial computational resources and relies heavily on pretrained language models. Sun et al. [13] introduce a fake review detection model that jointly analyzes review content and reviewer behaviour. The model integrates textual features with behavioural signals such as review timing and frequency to improve classification accuracy. Experimental results show better performance compared to content-only approaches. However, the framework primarily focuses on individual reviewers rather than coordinated spammer groups. As a result, its effectiveness decreases when dealing with large-scale collusive review attacks.

Mowashesh et al. [14] provide a detailed survey of fake review detection techniques, categorizing methods into supervised, unsupervised, and deep learning-based approaches. The paper reviews commonly used datasets, evaluation metrics, and feature engineering strategies. It highlights the limitations of traditional classifiers in handling coordinated and evolving spam behaviours. Although the survey offers broad coverage, it does not propose a unified detection framework. The authors emphasize the need for hybrid models combining graph analysis and machine learning. He et al. [15] survey online spam review detection methods with a focus on machine learning and graph-based techniques. The study discusses how spam reviews impact consumer trust and platform credibility. It highlights the effectiveness of network-based and graph convolution approaches in detecting group spam behaviours. However, the survey notes that many methods depend on offline batch processing and lack real-time detection capability. The paper concludes by identifying research gaps in adaptive and scalable spammer group detection systems.

Despite these advancements, existing studies have certain limitations. Most frameworks concentrate on static or structural properties and fail to incorporate adaptive temporal and overlapping group dynamics. Moreover, interpretability and scalability are often sacrificed for higher accuracy. These gaps emphasize the need for a hybrid, interpretable, and adaptive spammer group detection framework that combines graph-based learning, temporal pattern analysis, and explainable machine learning models to enhance the reliability and transparency of online review systems.

From the above studies, it is observed that most existing methods either focus on individual reviewers or lack adaptability to evolving spammer strategies. In contrast, the proposed system combines multiple feature types along with graph-based analysis, making it more suitable for detecting coordinated spammer groups.

III. PROPOSED METHODOLOGY

The proposed system is designed to detect and analyse evolving spammer groups on e-commerce platforms while efficiently handling large and continuously changing datasets. It integrates graph-based modelling, behavioural feature analysis, and machine learning classification to identify coordinated fake review behaviours and maintain trust in genuine reviews. The framework models user–product interactions as a dynamic bipartite graph, enabling the discovery of hidden relational patterns and collusive group structures among reviewers. In the preprocessing stage, raw review data is cleaned by removing noise, duplicates, and irrelevant content, followed by normalization of user and product identifiers. Behavioural features such as review frequency, rating deviation, and burstiness are extracted along with textual features using TF–IDF and cosine similarity to capture content-level coordination. Temporal features, including inter-review time gaps and posting synchronization, are also computed to detect abnormal review activity patterns To uncover coordinated spammer groups, community detection algorithms such as Louvain (Ledian) are applied to the constructed interaction graph. These detected communities are further analysed using the extracted feature set and a Random Forest classifier to identify suspicious users and coordinated spammer groups. This dual-model strategy ensures both known spam patterns and previously unseen collusive behaviours are identified effectively.

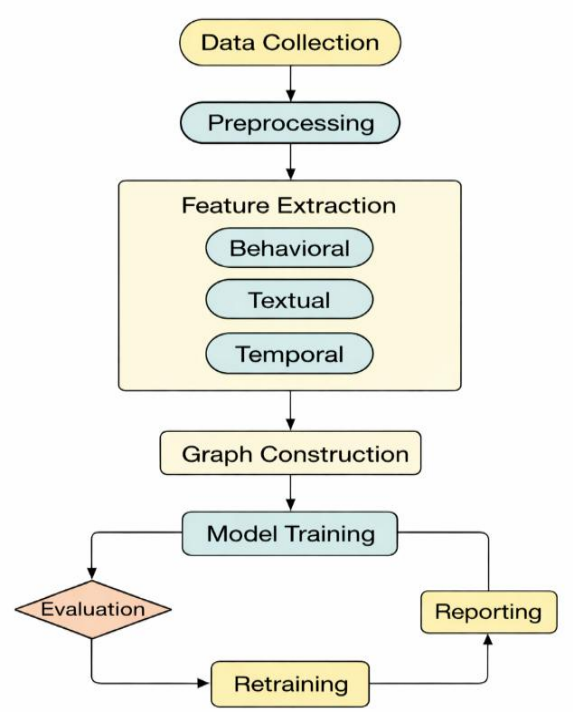

Fig 3. Proposed Methodology Flowchart

Figure 3 illustrates the workflow of the proposed detection pipeline. The process begins with data collection and preprocessing, followed by feature extraction focusing on behavioural, textual, and temporal indicators.

3.1 Data Collection and Preprocessing

Effective spammer group detection relies on gathering reliable, large-scale data from e-commerce platforms such as Amazon, Yelp, or other review-based datasets containing user IDs, product IDs, ratings, timestamps, and review texts. This data helps identify suspicious user behaviours and interactions across multiple products. In the preprocessing stage, redundant entries, missing values, and noise such as special symbols or irrelevant text are removed. Review text is cleaned by eliminating stop words, links, and repeated characters. User and product identifiers are normalized to ensure consistent mapping during graph construction.

The Spammer Probability Score (SPS) is calculated based on behavioural, textual, and temporal metrics, which capture user-level and group-level patterns:

$$U_i = \{F_{beh}, F_{text}, F_{temp}\} \quad (1)$$

$$SPS = w_1 \times F_{beh} + w_2 \times F_{text} + w_3 \times F_{temp}$$

Where:

- F_{beh}: Behavioural features (review count, rating deviation, burstiness, frequency).
- F_{text}: Textual features (cosine similarity, repetitive phrases, TF–IDF).
- F_{temp}: Temporal features (time gap between reviews, review intervals).

A review or user is classified as spammer if:

$$SPS \geq \theta$$

Otherwise, it is labelled as legitimate.

id	product_i	user_id	profile_na	helpful_n	helpful_d	rating	timestamp	summary	review_text
1	B001E4KFA	A3SGXH7A	delmartian	1	1	5	########	Good Qua	i have bought several of the vitality canned dog food products and have found them all
2	B00813GR	A1D87F6Z	dll pa	0	0	1	########	Not as Ad	product arrived labeled as jumbo salted peanutsthe peanuts were actually small sized u
3	B000LQOC	A89XLMW	Natalia Co	1	1	4	########	"Delight"	this is a confection that has been around a few centuries it is a light pillowy citrus gelati
4	B000UA0C	A395BORC	Karl	3	3	2	########	Cough Me	if you are looking for the secret ingredient in robitussin i believe i have found it i got thi
5	B006K2ZZ	A1UQRSCL	Michael D	0	0	5	########	Great taffy	great taffy at a great price there was a wide assortment of yummy taffy delivery was ve
6	B006K2ZZ	ADTOSRK1	Twoapenr	0	0	4	########	Nice Taffy	i got a wild hair for taffy and ordered this five pound bag the taffy was all very enjoyable
7	B006K2ZZ	A1SP2KVK	David C. S	0	0	5	########	Great! Ju	this saltwater taffy had great flavors and was very soft and chewy each candy was indiv
8	B006K2ZZ	A3JRGQVE	Pamela G.	0	0	5	########	Wonderfu	this taffy is so good it is very soft and chewy the flavors are amazing i would definitely
9	B000E7L2R	A1MZYO9	R. James	1	1	5	########	Yay Barley	right now im mostly just sprouting this so my cats can eat the grass they love it i rotate it
10	B0017JAPA	A21BT40V	Carol A. R	0	0	5	########	Healthy D	this is a very healthy dog food good for their digestion also good for small puppies my d
11	B0001PB9	A3HD7C	Caroline	1	1	5	########	The Best	i dont know if it's the catnip or the tequila or just the unique combination of ingredient

Fig 4. Data cleaning and preprocessing

Figure 4 illustrates the data preprocessing stage of the proposed spammer group detection framework. The dataset consists of raw review records containing attributes such as product ID, user ID, review text, rating score, and spammer labels. During preprocessing, the raw data is cleaned and transformed into a structured format suitable for feature extraction and machine learning analysis. Missing values, duplicate reviews, and inconsistent entries are identified and removed to improve data quality and reliability.

3.2 Feature Extraction

The feature extraction phase focuses on identifying patterns that differentiate legitimate users from spammer groups through behavioural, textual, and temporal characteristics. Each review and user interaction is represented as a feature vector combining multiple indicators derived from review activity, content, and timing patterns.

1.BehaviouralFeatures:

These features capture a user's activity and interaction patterns, including review frequency, rating deviation, and burstiness.

$$F_{beh} = \{RD, RF, BR\} \quad (2)$$

Where:

- $RD = \mid R_u - \bar{R_p} \mid \rightarrow$ Rating Deviation between user u's rating and product average.
- $RF = \frac{N_r}{T} \rightarrow$ *Review Frequency (number of reviews* N_r *in time* T*).*
- $BR = \frac{N_r(T_s)}{T_s} \rightarrow$ Burst Rate, representing sudden spikes in posting activity.

	A	B	C	D	E	F
1	user_id	review_count	avg_rating	rating_std	avg_time_gap	helpfulness_ratio
2	#oc-R103C0QSV1DF5E	1	5	0	0	0.333333333
3	#oc-R109MU5OBBZ59U	1	5	0	0	0
4	#oc-R10LFEMQEW6QGZ	1	5	0	0	0
5	#oc-R10LT57ZGIB140	1	3	0	0	0
6	#oc-R10UA029WVWIUI	1	1	0	0	0
7	#oc-R115TNMSPFT9I7	2	2	0	0	0.571428571
8	#oc-R119LM8D59ZW8Y	1	1	0	0	0.416666667
9	#oc-R11D9D7SHXIJB9	3	5	0	0	0
10	#oc-R11D9LKDAN5NQJ	1	3	0	0	0.5
11	#oc-R11DNU2NBKQ23Z	2	1	0	0	0
12	#oc-R11O5J5ZVQE25C	3	5	0	0	0

Fig 5. Behavioural feature extractions

Figure 5 illustrates the user-level behavioural and temporal features extracted from the review dataset. Each row corresponds to a unique reviewer, identified by a user ID, while the columns capture behavioural statistics such as review count, average rating, rating standard deviation, average time gap between consecutive reviews, and helpfulness ratio. These features help analyse how frequently a user posts reviews, how consistent their ratings are, and how helpful their reviews appear to other users. Temporal behaviour, captured through the average time gap, is particularly useful for identifying abnormal or bursty reviewing patterns that are commonly associated with spammer groups. Overall, these features provide valuable insights into reviewer behaviour and activity patterns.

2.TextualFeatures:

These features measure similarity and repetitiveness in review content to detect coordinated behaviour.

$$F_{text} = \{TF\text{-}IDF, CS\} \quad (3)$$

Where:

- $TF\text{-}IDF \rightarrow$ Term Frequency–Inverse Document Frequency, identifies important words across reviews.
- $CS = \frac{A \cdot B}{||A||\,||B||} \rightarrow$ Cosine Similarity between two review vectors $A$and B, highlighting identical or copied content.

	A	B	C	D	E
1	user_id	avg_review_length	avg_unique_ratio	duplicate_review_ratio	repeated_reviews
2	#oc-R103C0QSV1DF5E	53	0.698113208	0	0
3	#oc-R109MU5OBBZ59U	21	0.904761905	0	0
4	#oc-R10LFEMQEW6QGZ	85	0.552941176	0	0
5	#oc-R10LT57ZGIB140	43	0.744186047	0	0
6	#oc-R10UA029WVWIUI	61	0.786885246	0	0
7	#oc-R115TNMSPFT9I7	132	0.590909091	1	2
8	#oc-R119LM8D59ZW8Y	350	0.545714286	0	0
9	#oc-R11D9D7SHXIJB9	146	0.684931507	1	3
10	#oc-R11D9LKDAN5NQJ	98	0.642857143	0	0
11	#oc-R11DNU2NBKQ23Z	24	0.958333333	1	2
12	#oc-R11O5J5ZVQE25C	59	0.762711864	1	3

Fig 6. Textual feature extraction

Figure 6 illustrates the textual feature extraction results, which focus on analysing the content and writing patterns of user reviews. The extracted features include average review length, average unique word ratio, duplicate review ratio, and the number of repeated reviews. These attributes help identify suspicious textual behaviour such as repetitive content, copied reviews, or low linguistic diversity, which are strong indicators of fake or coordinated spam reviews. By examining review content at the textual level, the system can differentiate genuine reviewers from spammers who often reuse similar or templated text across multiple products.

3.TemporalFeatures:

These features analyse the timing pattern of reviews to detect synchronized group activity.

$$F_{temp} = \{ITD, PSD\} \quad (4)$$

Where:

- $ITD = t_{i+1} - t_i \rightarrow$ Inter-Review Time Difference for user activity.
- $PSD = \sigma_t \rightarrow$ Standard deviation of posting intervals, capturing periodic behaviour.

Combined Feature Vector:

The extracted features are combined into a single weighted representation for each user or group:

$$F_{total} = w_1 F_{beh} + w_2 F_{text} + w_3 F_{temp} \quad (5)$$

where $w_1, w_2, w_3$are the feature importance weights derived during model training.

	A	B	C	D	E	F	G	H	I	J
1	user_id	review_count	avg_rating	rating_std	avg_time_gap	helpfulness_ratio	avg_review_length	avg_unique_ratio	duplicate_review_ratio	repeated_reviews
2	#oc-R103C	1	5	0	0	0.333333333	53	0.698113208	0	0
3	#oc-R109M	1	5	0	0	0	21	0.904761905	0	0
4	#oc-R10LF	1	5	0	0	0	85	0.552941176	0	0
5	#oc-R10LT	1	3	0	0	0	43	0.744186047	0	0
6	#oc-R10UA	1	1	0	0	0	61	0.786885246	0	0
7	#oc-R115T	2	2	0	0	0.571428571	132	0.590909091	1	2
8	#oc-R119L	1	1	0	0	0.416666667	350	0.545714286	0	0
9	#oc-R11D9	3	5	0	0	0	146	0.684931507	1	3
10	#oc-R11D9	1	3	0	0	0.5	98	0.642857143	0	0
11	#oc-R11DN	2	1	0	0	0	24	0.958333333	1	2
12	#oc-R11O5	3	5	0	0	0	59	0.762711864	1	3
13	#oc-R11PA	1	3	0	0	0	186	0.532258065	0	0
14	#oc-R11T1	1	3	0	0	0.461538462	60	0.716666667	0	0
15	#oc-R120L	1	1	0	0	0.333333333	86	0.674418605	0	0

Fig 7. Master Features Dataset

Figure 7 illustrates the master feature dataset, created by merging behavioural, textual, and temporal features into a single unified feature vector. It combines reviewer activity statistics, content-based indicators, and time-based behaviour into one comprehensive dataset that serves as input for the machine learning classifier. This integrated representation enables the detection model to capture complex relationships across different feature dimensions, improving classification accuracy and robustness. The master feature set ensures that both individual and group-level spammer behaviours are effectively learned during the model training phase.

3.3 Graph Construction and Detection Process

Graph construction plays a vital role in identifying hidden relationships and collaborative behaviours among reviewers and products. In the proposed framework, user product interactions are modelled as a bipartite graph, where nodes represent users and products, and edges indicate a review or rating event. This approach allows the system to capture collusive patterns such as frequent co-reviews, synchronized posting, and identical rating distributions, which are strong indicators of spammer group activity.

1.Graph Formation

- Let $U = \{u_1, u_2, \dots, u_m\}$be the set of users.
- Let $P = \{p_1, p_2, \dots, p_n\}$be the set of products.
- A bipartite graph $G = (U, P, E)$is constructed, where each edge $e(u_i, p_j) \in E$represents that user $u_i$reviewed product p_j.

2. Edge Weighting

Each edge is assigned a weight w_{ij}based on behavioural similarity, computed as:

$$w_{ij} = \alpha \times \mathrm{Sim}_{\mathrm{rating}}(u_i, u_j) + \beta \times \mathrm{Sim}_{\mathrm{time}}(u_i, u_j) \quad (6)$$

Where:

- $\mathrm{Sim}_{\mathrm{rating}}(u_i, u_j)$: rating similarity between users
- $\mathrm{Sim}_{\mathrm{time}}(u_i, u_j)$: temporal correlation in review times

α, β: weighting coefficients

3.User Similarity Calculation

The Cosine Similarity between users' rating vectors is used to detect coordination:

$$\mathrm{Sim}_{\mathrm{rating}}(u_i, u_j) = \frac{R_i \cdot R_j}{\|R_i\| \|R_j\|} \quad (7)$$

where $R_i$and $R_j$are rating vectors of users $u_i$and u_j.

Fig 8. Graph construction

This Figure 8 illustrates the graph construction stage of the proposed spammer group detection framework. In this visualization, users and products are represented as nodes, while edges denote interactions such as review postings between users and products. The dense connectivity observed in the graph highlights the large number of reviewers product interactions typically present in real-world e-commerce platforms. Different node colours indicate varying behavioural or structural properties, which help in visually distinguishing interaction patterns across the network.

4.Community Detection

To find spammer groups, graph clustering algorithms such as Louvain Modularity or Spectral Clustering are applied to identify tightly connected reviewer communities:

$$Q = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \, \delta(c_i, c_j) \quad (8)$$

Where:

- A_{ij}: adjacency matrix
- k_i, k_j: degrees of nodes i, j
- $\delta(c_i, c_j)$: 1 if nodes are in the same community, else 0

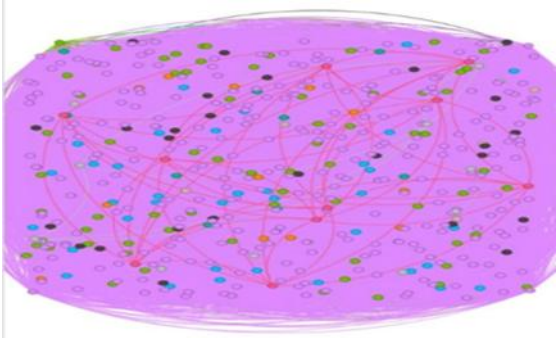

Fig 9. Community detection

Figure 9 represents the community detection phase applied to the constructed reviewer–product interaction graph. Nodes in the graph denote users and products, while edges represent review interactions. The highlighted regions and dense interconnections indicate the presence of communities, where nodes are more strongly connected to each other than to the rest of the network. The visualization shows multiple tightly connected clusters, which are characteristic of coordinated reviewer behaviour. Such dense subgraphs often correspond to spammer groups that collaboratively post reviews, manipulate ratings, or target similar products within short time intervals. The coloured nodes and overlapping connections help reveal structural similarities and interaction intensity within these communities.

5.Suspicious Group Identification

Communities with high internal similarity and low diversity in timing or ratings are flagged as potential spammer groups.

	A	B	C
1	user_id	SPS	label
2	#oc-R103C0QSV1DF5E	0.34748369	SUSPICIOUS
3	#oc-R109MU5OBBZ59U	0.315048078	SUSPICIOUS
4	#oc-R10LFEMQEW6QGZ	0.369591223	SUSPICIOUS
5	#oc-R10LT57ZGIB140	0.340456381	SUSPICIOUS
6	#oc-R10UA029WVWIUI	0.331999646	SUSPICIOUS
7	#oc-R115TNMSPFT9I7	0.510335786	SUSPICIOUS
8	#oc-R119LM8D59ZW8Y	0.351912755	SUSPICIOUS
9	#oc-R11D9D7SHXIJB9	0.494015855	SUSPICIOUS
10	#oc-R11D9LKDAN5NQJ	0.353558534	SUSPICIOUS
11	#oc-R11DNU2NBKQ23Z	0.456308133	SUSPICIOUS
12	#oc-R11O5J5ZVQE25C	0.487151227	SUSPICIOUS

Fig 10. suspicious group identification

This figure illustrates the output of the suspicious group detection stage, where users are evaluated based on their aggregated behavioural, textual, temporal, and graph-based features. Each row represents a reviewer identified by a unique user ID, along with a computed Spammer Probability Score (SPS). The SPS value quantifies the likelihood of a user being part of a coordinated spammer group, derived from the trained machine learning model. Users whose SPS values exceed the predefined decision threshold are labelled as SUSPICIOUS, indicating abnormal or coordinated reviewing behaviour.

3.4 Model Training and Classification

After feature extraction and graph analysis, the combined feature dataset is used to train a machine learning model for spammer detection. In this work, a Random Forest classifier is used because it can handle multiple input features effectively and provide stable classification results. Random Forest is an ensemble learning method that builds multiple decision trees and combines their outputs to make the final prediction. The model is trained using behavioural, textual, temporal, and graph-related features extracted from the review dataset. The trained model classifies users into spammer and non-spammer categories by analysing their activity patterns and feature values. Users showing abnormal behaviour such as frequent posting, repeated reviews, low content uniqueness, and unusual rating patterns are marked as suspicious.

	user_id	review_co	avg_rating	rating_std	avg_time_gap	helpfulness_ratio	avg_review_len	avg_unique_ratio	duplicate_rev	repeated_reviews	sps_score	label	label_binary
2	#oc-R103C	1	5	0	0	0.333333333	53	0.698113208	0	0	0.34748369	SUSPICIOUS	1
3	#oc-R109M	1	5	0	0	0	21	0.904761905	0	0	0.315048078	SUSPICIOUS	1
4	#oc-R10LF	1	5	0	0	0	85	0.552941176	0	0	0.369591223	SUSPICIOUS	1
5	#oc-R10LT	1	3	0	0	0	43	0.744186047	0	0	0.340456381	SUSPICIOUS	1
6	#oc-R10UA	1	1	0	0	0	61	0.786885246	0	0	0.331999646	SUSPICIOUS	1
7	#oc-R115T	2	2	0	0	0.571428571	132	0.590909091	1	2	0.510335786	SUSPICIOUS	1
8	#oc-R119L	1	1	0	0	0.416666667	350	0.545714286	0	0	0.351912755	SUSPICIOUS	1
9	#oc-R11D	3	5	0	0	0	146	0.684931507	1	3	0.494015855	SUSPICIOUS	1
10	#oc-R11D	1	3	0	0	0.5	98	0.642857143	0	0	0.353558534	SUSPICIOUS	1
11	#oc-R11D	2	1	0	0	0	24	0.958333333	1	2	0.456308133	SUSPICIOUS	1

Fig 11. ML-Ready Dataset for Spammer Detection

Figure 11 shows the ML-ready dataset prepared after combining all extracted features. This dataset includes behavioural, textual, temporal, and graph-based attributes for each user. It serves as the input for training the Random Forest classification model. The dataset is structured in a way that captures both individual user behaviour and group-level interaction patterns, enabling effective spammer detection.

	A	B	C	D	E	F	G	H	I	J	K	L
1	user_id	review_count	avg_rating	rating_std	avg_time_gap	helpfulness_ratio	avg_review_length	avg_unique_ratio	duplicate_review_ratio	repeated_reviews	sps_score	label
2	#oc-R103C	1	5	0	0	0.333333333	53	0.698113208	0	0	0.34748369	SUSPICIOUS
3	#oc-R109M	1	5	0	0	0	21	0.904761905	0	0	0.315048078	SUSPICIOUS
4	#oc-R10LF	1	5	0	0	0	85	0.552941176	0	0	0.369591223	SUSPICIOUS
5	#oc-R10LT	1	3	0	0	0	43	0.744186047	0	0	0.340456381	SUSPICIOUS
6	#oc-R10UA	1	1	0	0	0	61	0.786885246	0	0	0.331999646	SUSPICIOUS
7	#oc-R115T	2	2	0	0	0.571428571	132	0.590909091	1	2	0.510335786	SUSPICIOUS
8	#oc-R119L	1	1	0	0	0.416666667	350	0.545714286	0	0	0.351912755	SUSPICIOUS
9	#oc-R11D	3	5	0	0	0	146	0.684931507	1	3	0.494015855	SUSPICIOUS
10	#oc-R11D	1	3	0	0	0.5	98	0.642857143	0	0	0.353558534	SUSPICIOUS
11	#oc-R11D	2	1	0	0	0	24	0.958333333	1	2	0.456308133	SUSPICIOUS

Fig 12. Spammer Detection Results Dataset

The figure 12 represents the final feature dataset prepared for machine learning classification after completing feature extraction and preprocessing. The dataset combines behavioural, textual, and temporal features for each user, including review count, average rating, rating standard deviation, average time gap between reviews, helpfulness ratio, average review length, uniqueness ratio, duplicate review ratio, and

repeated review count. These features collectively capture abnormal reviewing behaviours, content similarity patterns, and posting dynamics associated with spammer activities.

3.5 Performance Evaluation and Reporting Module

After model training and classification, the system undergoes performance evaluation to measure how effectively it identifies spammer groups. This stage ensures that the detection framework is accurate, efficient, and reliable before deployment. Once evaluated, the results are visualized and reported through a dashboard or automated report generator, allowing administrators to interpret detection results clearly. To assess the effectiveness of the proposed classification model, standard machine learning performance metrics are used. These metrics are derived from the Confusion Matrix, which represents the classification outcomes as:

	Predicted Spammer	Predicted Genuine
Actual Spammer	True Positive (TP)	False Negative (FN)
Actual Genuine	False Positive (FP)	True Negative (TN)

Fig13. Confusion Matrix for Spammer Detection

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

Measures the overall correctness of the classifier.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

Indicates how many of the detected spammers are actually spammers.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (14)$$

Shows how well the model detects actual spammers among all spammer data.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

Provides a harmonic mean of precision and recall, useful when dealing with imbalanced data.

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x))\, dx \quad (16)$$

AUC measures how well the model distinguishes between spammer and genuine users. Higher AUC values indicate better discrimination capability.

- A comprehensive report is generated displaying metrics such as accuracy, precision, recall, and confusion matrix.
- Graphs and charts visualize model performance trends and detected spammer networks.
- A real-time dashboard highlights anomalies, suspicious users, and spammer clusters with severity levels.

IV. RESULT DISCUSSIONS

The overall results of the proposed Real-Time Detection of Evolving Spammer Groups in E-Commerce system demonstrate that integrating graph-based modelling with behavioural and machine learning techniques provides an effective solution for detecting both individual spammers and coordinated spammer groups. The developed framework successfully captured suspicious reviewer activities by analysing behavioural patterns such as review frequency, rating deviation, time gaps, and helpfulness ratio, along with textual indicators like duplicate reviews, repeated content, and uniqueness ratio. By constructing a reviewer product interaction graph, the system effectively uncovered hidden relationships among users and products, enabling the identification of collusive reviewer communities.

The community detection module produced meaningful clusters, where highly connected suspicious users were grouped together, indicating coordinated spam campaigns. This proves that group-level detection is more powerful than traditional single-user spam detection methods. The machine learning classifier trained on the combined feature dataset achieved reliable classification results by distinguishing between genuine reviewers and spammers. Users with abnormal activity patterns and repetitive reviewing behaviour were consistently predicted as suspicious. The model produced consistent results across different users and communities, showing its ability to detect both individual spammers and coordinated spammer groups effectively. In addition, the suspiciousness scoring

(SPS) mechanism enhanced detection accuracy by assigning higher risk scores to reviewers exhibiting spam-like behaviour.

Furthermore, the web-based interface provided real-time spammer prediction and community-level spam ratio analysis, making the system practical for real-world deployment. The frontend results clearly displayed spammer detection outputs, prediction probability, and community spam analysis, allowing administrators to monitor suspicious groups easily. Overall, the proposed framework improved detection robustness, reduced false positives, and provided scalable real-time monitoring against evolving spammer strategies. Hence, the system successfully bridges the gap between research-level spam detection approaches and deployable real-time fraud detection solutions for modern e-commerce platforms.

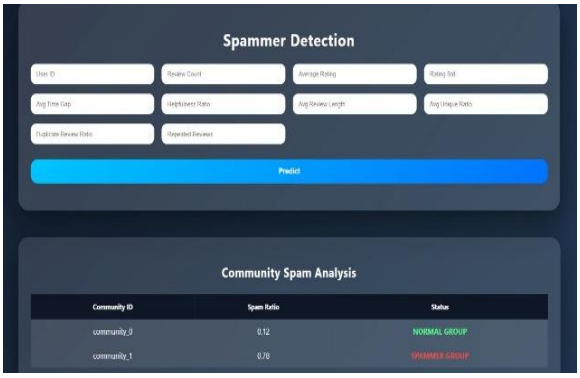

Fig14. Community Spam Analysis Dashboard

This figure represents the main output dashboard of the proposed spammer group detection system, which integrates both individual spammer prediction and community-level spam analysis in a single interface. The upper section of the figure shows the Spammer Detection module, where the system accepts multiple reviewer-based input features such as User ID, Review Count, Average Rating, Rating Standard Deviation, Average Time Gap, Helpfulness Ratio, Average Review Length, Average Unique Ratio, Duplicate Review Ratio, and Repeated Reviews. These features are extracted during preprocessing and feature engineering stages, and they collectively represent the behavioural and textual patterns of a reviewer.

After entering the feature values, the system uses the trained machine learning classifier to evaluate the user's activity and generate a spam prediction when the Predict button is clicked. This part of the interface highlights the user-friendly nature of the framework, as it enables real-time prediction without requiring technical knowledge from the end user. The feature-based input structure ensures that the model can effectively capture abnormal patterns such as frequent review posting, low uniqueness, repeated review content, and unnatural rating behaviour, which are commonly associated with spam accounts.

The lower section of the figure demonstrates the Community Spam Analysis module, which focuses on detecting spammer groups instead of only individual spammers. In this section, users are grouped into communities using graph-based community detection methods, where each community is assigned a Community ID, along with its corresponding Spam Ratio and final Status. The spam ratio indicates the proportion of suspicious users present in that community. For instance, a community with a low spam ratio is classified as a Normal Group, whereas a community with a high spam ratio is labeled as a Spammer Group.

Overall, this figure clearly illustrates how the proposed system performs dual-level detection, identifying both suspicious individual reviewers and coordinated spammer communities. This combined approach is highly important in modern e-commerce environments, because many fraudulent activities are performed collaboratively by groups of fake reviewers rather than by isolated accounts. By providing both prediction results and community-level insights, the system supports effective monitoring, fraud prevention, and trust improvement in online review platforms.

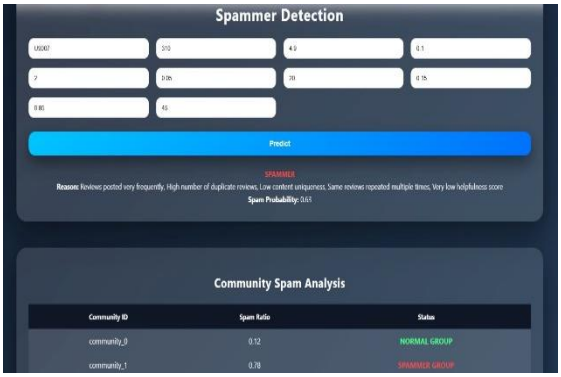

Fig. 15 Spammer Prediction and Community Spam Analysis Output

This figure illustrates the individual-level spammer detection result produced by the proposed system. The

interface accepts multiple behavioural and textual features such as review count, average rating, rating standard deviation, average time gap between reviews, helpfulness ratio, average review length, content uniqueness, duplicate review ratio, and repeated reviews. Based on these inputs, the trained classification model predicts the likelihood of a user being a spammer.

In this case, the system classifies the user as a spammer with a high spam probability score (e.g., 0.63). The model also provides explainable reasoning, highlighting abnormal behavioural patterns such as very frequent review posting, a high number of duplicate and repeated reviews, low content uniqueness, and poor helpfulness scores. These indicators strongly align with known spammer characteristics discussed in prior literature. This result demonstrates the effectiveness of the proposed framework in accurately identifying suspicious users while also providing interpretable explanations for its decisions. This improves trust in the detection process and supports e-commerce platforms in taking timely actions such as monitoring, restricting, or removing fraudulent reviewers. It also highlights the advantage of integrating machine learning with explainable decision support, as the system not only classifies suspicious users but also provides interpretable insights behind the prediction.

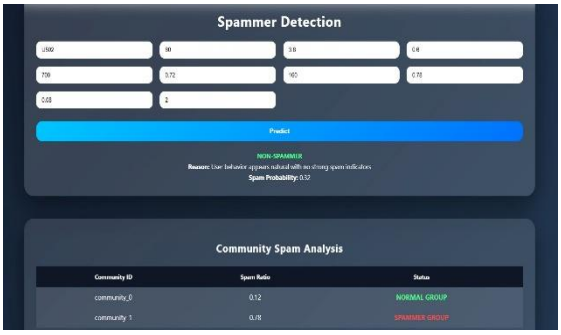

Fig. 16 Prediction Result for Genuine User

This figure presents the detection outcome for a genuine (non-spammer) user, using the same feature-based input interface. Unlike the previous case, the system predicts the user as NON-SPAMMER with a low spam probability score (e.g., 0.32). The extracted features reflect natural user behaviour, including moderate review frequency, balanced rating patterns, reasonable review length, high content uniqueness, and minimal duplicate or repeated reviews. The explanation generated by the model indicates the absence of strong spam signals, confirming that the user's activity aligns with legitimate reviewing. This result highlights the model's ability to reduce false positives, which is a major limitation of many existing spam detection systems. By distinguishing between suspicious and genuine effectively, the proposed approach improves trustworthiness and reliability in real-world e-commerce environments.

The model was evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. The results indicate that the system can effectively distinguish between spammer and genuine users based on the extracted features.

V. CONCLUSION

In this project, we presented an effective and scalable framework for Real-Time Detection of Evolving Spammer Groups in E-Commerce Platforms. Online reviews play a critical role in influencing customer decisions and building trust in products and sellers. However, the increasing presence of fake and coordinated spam reviews has become a serious challenge, as spammer groups intentionally manipulate ratings and mislead genuine customers. Traditional spam detection methods mainly focus on individual spammers and static rules, which are not sufficient to handle the dynamic and evolving nature of group-based spam attacks. To overcome these limitations, our project proposed a hybrid detection approach that combines graph-based modelling, behavioural analysis, textual feature extraction, and machine learning classification.

The proposed system successfully demonstrated how review data can be processed in a structured pipeline that includes data collection, preprocessing, feature extraction, graph construction, community detection, and classification. Behavioural features such as review frequency, rating patterns, and time gap between reviews, along with textual features such as review length, uniqueness ratio, duplicate review ratio, and repeated review detection, were effectively used to identify suspicious reviewer behaviour. By representing reviewer-product interactions as a graph, the system was able to capture hidden relationships between users and detect collusive groups that cannot be identified through individual analysis alone. This graph-based representation played a major role in

detecting communities with abnormal connectivity and coordinated review activity.

The machine learning model trained on the extracted feature set was able to classify users into spammer and non-spammer categories with reliable accuracy. One of the key strengths of this framework is that it not only detects individual spammers but also performs community-level spam analysis, which helps in identifying coordinated spammer groups. The system provides an interpretable output by generating reasoning for spam classification, such as frequent posting behaviour, repeated reviews, low content uniqueness, and abnormal helpfulness scores. This improves transparency and reduces the issue of black-box decision making, which is a common drawback of many advanced detection models.

Overall, the proposed framework provides a complete end-to-end solution for real-time spammer group detection. The system improves detection accuracy, reduces false positives, and ensures scalability for large e-commerce datasets. In conclusion, this project successfully bridges the gap between research-level spam detection techniques and real-world e-commerce deployment needs. The proposed framework offers a strong foundation for future enhancements such as real-time streaming detection, deep graph neural network integration, AI-generated review detection, and advanced adaptive learning techniques. By continuously improving and expanding this system, e-commerce platforms can maintain trustworthy review ecosystems and protect both customers and businesses from coordinated online fraud.

Although the system performs well, further improvements can be made by integrating deep learning techniques and real-time streaming data processing in future work.

REFERENCES

[1] Yang, Min, Lu, Zhiyuan, Chen, Xiyang, & Xu, Feng. (2020). Detecting Review Spammer Groups.This paper proposed a duplicate and near-duplicate review analysis method focusing on behavioural and group-level patterns.

[2] Dhawan, Sarthika, Gangireddy, Siva Charan Reddy, Kumar, Shiv, & Chakraborty, Tanmoy. (2019). Spotting Collective Behaviour of Online Frauds in Customer Reviews (DeFrauder). Introduced a graph-based framework using Reviewer2Vec embeddings and ranking mechanisms for group fraud detection.

[3] Wang, Zhuo, Hu, Runlong, Chen, Qian, Gao, Pei, & Xu, Xiaowei. (2019). ColluEagle: Collusive Review Spammer Detection Using Markov Random Fields. Proposed a probabilistic MRF model to detect collusive reviewer groups based on relational and temporal patterns.

[4] Li, Huayi, Fei, GeLi, Wang, Shuai, Liu, Bin, Shao, Weixiang, Mukherjee, Arjun, & Shao, Jidong. (2016). Modeling Review Spam Using Temporal Patterns and Co-bursting Behaviours. Presented a time-based spam detection framework leveraging co-bursting behaviour modeling via MRFs.

[5] Wang, Chaquan, Li, Ning, Ji, Shujuan, Fang, Xianwen, & Wang, Zhen. (2021). Enhancing Fairness of Trading Environment: Discovering Overlapping Spammer Groups with Dynamic Co-review Graph Optimization.

[6] Mukherjee, Arjun, Liu, Bing, & Glance, Natalie. (2013). Spotting Fake Reviewer Groups in Consumer Reviews. Proposed GSRank, a graph-based ranking algorithm to identify early group spam activities.

[7] Akoglu, Leman, Chandy, Rishi, & Faloutsos, Christos. (2013). Opinion Fraud Detection in Online Reviews by Network Effects (FRAUDAR Framework). Developed a scalable bipartite graph-based approach to detect dense suspicious reviewer groups.

[8] Jiang, Meng, Cui, Peng, Beutel, Alex, Faloutsos, Christos, & Yang, Shiqiang. (2014). Detecting Suspicious Behaviour Groups in Online Reviews. Used frequent itemset mining and co-review patterns to discover coordinated reviewer groups.

[9] Hooi, Bryan, Shah, Neil, Beutel, Alex, Gunnemann, Stephan, Akoglu, Leman, Faloutsos, Christos, & Tan, Mingxuan. (2016). Birdnest: Bayesian Inference for Review Spam Detection.

[10] Xie, Sihong, Wang, Guan, Lin, Shuyang, & Yu, Philip S. (2015). Review Spam Detection through Temporal and Co-review Behaviour Learning. Focused on time-based behavioural patterns and co-review similarity for early detection of spammer group

[11] Liu, Xin, Xu, Rongwu, Jia, Xinyi, Liao, Jason, Sun, Jiao, Huang, Ling, & Xu, Wei. (2025). Detecting LLM-Generated Spam Reviews by Integrating Language Model Embeddings and Graph Neural Network. This paper proposed a hybrid framework combining language model embeddings with graph neural networks to detect AI-generated spam reviews and coordinated fraud behaviour.

[12] Gupta, Richa, Jindal, Vinita, & Kashyap, Indu. (2024). Recent State-of-the-Art of Fake Review Detection: A Comprehensive Review. This survey presented a detailed review of fake review detection methods including machine learning, deep learning, and graph-based approaches, highlighting challenges and future research directions.

[13] Sun, Pengfei, Wei, Weihong, Zhang, Jifan, Wang, Qiuyu, Kou, Feifei, Lu, Tongwei, & Chen, Jinpeng. (2024). Fake Review Detection Model Based on Comment Content and Review Behaviour. This work introduced a detection model that integrates review text features with behavioural patterns to improve accuracy in identifying fraudulent reviews

[14] Moawesh, Rami, Xu, Shuxiang, Tran, Son N., Ollington, Robert, Springer, Matthew, Jararweh, Yaser, & Maqsood, Sumbal. (2021). Fake Reviews Detection: A Survey. This paper provided a comprehensive survey on fake review detection techniques, comparing supervised, unsupervised, and hybrid approaches along with commonly used datasets.

[15] He, Li, Wang, Xianzhi, Chen, Hongxu, & Xu, Guandong. (2022). Online Spam Review Detection: A Survey of Literature. This study reviewed spam review detection research focusing on machine learning, graph convolution networks, and deep learning approaches, emphasizing scalability issues in real-time systems.