AI-Driven Detection of Group Shilling and Review Analysis Using Bisecting K-Means

M. Revathi[1]; M. M. Harshitha[2]; Dr. Girish Kumar D.[3]

[1] PG Student, Department of MCA, Ballari Institute of Technology & Management, Ballari
[2]Assistant Professor, Department of MCA, Ballari Institute of Technology & Management, Ballari
[3] Professor and HoD, Department of MCA, Ballari Institute of Technology & Management, Ballari

Publication Date: 2026/05/12

Abstract: Online customer reviews strongly affect purchasing decisions on e-commerce platforms. However, the increasing presence of fake and biased reviews created through coordinated group shilling activities reduces the reliability of recommendation systems. This project presents an AI-based review analysis system designed to detect fraudulent and suspicious review behavior using NLP and ML techniques. The system adopts a role-based framework with separate access for users and administrators. Users submit product reviews and ratings, while administrators analyze the collected data. Review text is processed using NLP techniques, and Bisecting K-Means clustering is applied to identify groups of similar review patterns and user behaviors. Suspicious clusters indicating coordinated manipulation are detected, thereby improving review credibility and recommendation accuracy.

Keywords: *Artificial Intelligence, Natural Language Processing, Bisecting K-Means, Group Shilling Detection, Review Analysis, Machine Learning.*

How to Cite: M. Revathi; M. M. Harshitha; Dr. Girish Kumar D. (2026). AI-Driven Detection of Group Shilling and Review Analysis Using Bisecting K-Means. *International Journal of Innovative Science and Research Technology*, 11(4), 4210-4214. https://doi.org/10.38124/ijisrt/26apr680

I. INTRODUCTION

With the rapid growth of e-commerce platforms, online reviews have become an essential source of information for customers before making purchasing decisions. Reviews help users understand product quality, performance, and overall satisfaction. However, many online platforms face serious challenges due to fake and biased reviews generated intentionally to manipulate product ratings. These reviews are often created through group shilling attacks, where multiple users collaborate to promote or demote specific products.

To address this issue, intelligent systems volumes of reviews are required. This project focuses on detecting suspicious review behavior by combining. By using the Bisecting K-Means clustering algorithm, the system groups similar review patterns and identifies abnormal behavior, helping to improve the trustworthiness and effectiveness of online recommendation systems.

In this work, an intelligent review analysis framework is developed by integrating NLP-based preprocessing with the Bisecting K-Means clustering method to detect suspicious and coordinated review activities. The system is designed using a role-based access model that separates user and administrator functionalities. Users are allowed to submit product reviews and ratings, whereas administrators are responsible for examining review patterns, detecting abnormal clusters, and taking necessary actions.

Among the available clustering approaches, the Bisecting K-Means algorithm is particularly effective for identifying coordinated review manipulation, as it efficiently divides data into distinct and well-structured clusters.

II. RELATED WORK

Shyam Sundar Bhuvaneswari and Thangamuthu conducted a systematic review on intelligent safety systems and noted that the inclusion of IoT, sensor data, and integrated mapping APIs substantially improves the accuracy and speed of safety responses. Their work emphasized the growing importance of context-aware systems capable of adapting to dynamic threats through predictive detection and automated escalation mechanisms [1].

Dhanalakshmi et al. proposed a self-defense-driven mobile framework that combines user-triggered alerts with educational modules and law awareness. Their 2025 implementation demonstrated how blending GPS, video-based tutorials, and SMS-based fallback messaging can aid in ensuring accessibility even under network limitations [2].

Similarly, Chinnasamy et al. introduced real-time GPS to track users and alert law enforcement. Their system also

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr680

featured safe-route navigation and proximity alerts for high-risk areas, contributing to both reactive and preventive safety layers [3].

Chowdhuri's work on 'Suraksha', a Bluetooth-enabled self- defense kit, adds another dimension by fusing hardware (wearables) with app-based control. It incorporates physical triggers, such as hidden buttons in accessories, to activate alerts and deterrent sounds, addressing the usability barrier during active threats [4].

Lastly, Jeyanthi et al. introduced a quantum-assisted networked model to route alert data securely to law enforcement, tackling both latency and data tampering concerns. The fusion of legal knowledge modules with emergency functions was found to foster increased reporting rates and user confidence [5].

III. PROPOSED FRAMEWORK

➢ *System Overview*

The proposed framework presents an intelligent, AI-driven system designed to analyze online customer reviews and detect group shilling attacks using (NLP) and the Bisecting K-Means clustering algorithm. The framework aims to improve the reliability of online reviews by identifying fraudulent, biased, and coordinated review behaviors that can mislead consumers and distort recommendation systems.

The system is built on a role-based architecture consisting of two main modules: User Module and Admin Module. Each review is stored in the database with a unique user identifier and timestamp, ensuring proper tracking of user activity.

➢ *Key Functional Modules*

The proposed system consists of multiple functional components designed to manage user access, collect reviews, and store review-related data in a structured manner. A secure authentication mechanism handles user registration and login, ensuring role-based access for users and administrators. Authenticated users are allowed to submit product reviews and ratings, which are recorded along with user identifiers and timestamps in a centralized database. This structured data storage supports efficient review management and enables reliable tracking of user activity.

To prepare the collected reviews for analysis, the preprocessing stage includes text normalization, removal of irrelevant terms, and feature extraction to capture essential information from reviews. These processed features provide a suitable input format for ML algorithms, enabling accurate comparison and similarity measurement among reviews and user behaviors.

The core analysis component employs the Bisecting K-Means clustering algorithm to group reviews and users based on content similarity, rating behavior, and temporal patterns. By recursively dividing data into well-defined clusters, the system identifies abnormal groupings that may indicate coordinated shilling activities. Administrators can examine these suspicious clusters through an analysis interface, generate reports, and take appropriate actions to filter unreliable reviews. This modular approach enhances the reliability of online feedback and strengthens the credibility of recommendation systems.

This module generates summary reports based on the analysis results. It highlights suspicious users, detected shilling groups, and overall review statistics. The reports help administrators understand system behavior and take appropriate corrective actions to maintain review integrity.

➢ *Visual Overview*

The visual system illustrates the overall workflow and interaction between different modules involved in review collection and analysis. These reviews are stored in a centralized database that maintains user identifiers, review content, and submission time. This visual user-generated data flows securely from the front-end interface to the storage layer.

Finally, the visual overview demonstrates the administrative control and decision-making process. Based on the analysis, the admin can flag or filter unreliable reviews to maintain system integrity. This provides a clear data flow, module interaction, and analytical processing within the proposed framework.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr680

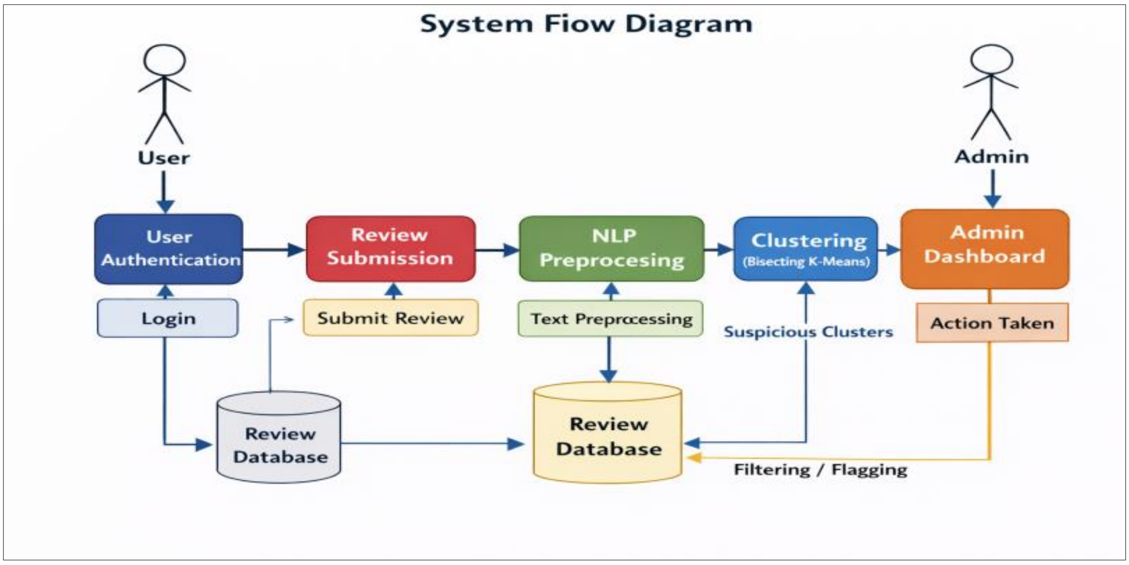

Fig 1 System Overflow

IV. METHODOLOGY AND IMPLEMENTATION

The methodology of the proposed system focuses on collecting user reviews, processing textual data, and identifying coordinated group shilling behavior using machine learning techniques. The implementation follows a systematic pipeline that ensures accurate review analysis and effective detection of fraudulent activities.

Initially, users register and log in to the system through a secure authentication mechanism. Authenticated users submit product reviews along with ratings, which are stored in a centralized database together with user identifiers and timestamps. This structured data collection enables proper tracking of user behavior and supports further analytical processing. The system ensures role-based access, allowing only administrators to view and analyze the complete review dataset.

In the next stage, the collected reviews undergo NLP to make data for ML analysis. These methods represent data numerically while capturing its key semantic aspects. The processed data is then used as input for the clustering phase. contacts or helpline numbers.

The core analytical stage applies the Bisecting K-Means clustering algorithm to group reviews and users based on similarities in content, rating patterns, and posting behavior. By recursively dividing the data into smaller and more homogeneous clusters, the algorithm helps uncover abnormal group patterns that indicate potential shilling attacks. Clusters showing unusually high similarity or synchronized behavior are marked as suspicious.

Finally, the admin dashboard, where administrators can inspect detected clusters and review detailed reports. Based on these insights, appropriate actions such as flagging or filtering unreliable reviews can be taken. This methodology ensures a reliable, scalable, and effective framework for enhancing the trustworthiness of online review systems.

V. RESULTS AND DISCUSSIONS

The proposed AI-driven using a dataset of user-submitted product reviews collected through the implemented web application. The dataset included reviews with varying ratings, textual content, and submission times from multiple users. The objective of the evaluation was to examine the system’s ability to organize reviews into meaningful clusters and identify suspicious group behaviors indicative of shilling attacks.

During preprocessing, the Natural Language Processing module successfully transformed unstructured review text into structured representations by removing Nonessential terms and extracting meaningful features. This preprocessing step reduced noise data and improved the similarity comparison between reviews. As a result, reviews with similar opinions, wording patterns, and sentiment characteristics were more accurately grouped during the clustering phase.

The Bisecting K-Means clustering algorithm produced well-defined clusters based on review content, rating behavior, and temporal patterns. Normal user behavior formed balanced clusters with diverse review content and varying submission times. In contrast, suspicious clusters were observed where multiple users posted highly similar reviews, identical rating values, or reviews within short time intervals. These clusters indicated coordinated behavior commonly associated with group shilling attacks.

The admin dashboard provided a clear visualization of clustered data, enabling administrators to examine suspicious groups and individual user activities. Through the analysis interface, administrators could easily identify abnormal patterns and flag unreliable reviews. The system demonstrated improved accuracy in detecting coordinated manipulation compared to manual inspection, reducing analysis time and effort.

The results also highlight the advantage of using

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr680

Bisecting K-Means over traditional flat clustering methods. The recursive splitting strategy enabled better separation of review groups, allowing subtle coordinated behaviors to be larger datasets. This scalable and effective for handling increasing volumes of user-generated content.

Overall, the results confirm that integrating NLP techniques with the Bisecting K-Means algorithm is effective for identifying fraudulent review behavior. The system enhances review reliability by filtering suspicious feedback and supports more accurate recommendation systems. further improvements can be achieved by incorporating larger datasets and advanced learning models for real-time analysis.

VI. CONCLUSION

The rapid growth of e-commerce platforms has made online reviews. However, the increasing presence of fake, biased, and manipulated reviews has reduced the reliability of review-based recommendation systems. This project addressed this challenge by proposing an AI-driven framework capable of analyzing online reviews and detecting coordinated group shilling activities using NLP and ML techniques.

The proposed system successfully integrates NLP-based preprocessing methods to handle unstructured textual review data. By cleaning, normalizing, and extracting meaningful features from reviews, the system improves the data used for analysis. This preprocessing stage ensuring accurate similarity measurement and effective clustering of review patterns.

The application of the Bisecting K-Means clustering algorithm proved effective in identifying hidden group behaviors within data. The recursive clustering approach enabled the system to separate normal user behavior from suspicious coordinated activities. Clusters exhibiting abnormal similarities in review content, rating patterns, and posting time were accurately identified as potential group shilling attacks.

The role-based architecture of the system further enhanced its effectiveness and usability. Users were provided simple and secure interface to submit reviews, while administrators were equipped with analytical tools to monitor review activity, inspect suspicious clusters, and take appropriate actions. This separation of roles ensured system security, transparency, and efficient review management.

FUTURE WORK

The proposed system lays a strong foundation for review analysis and management; to enhance its functionality and performance in the future. This would allow the system to better understand nuanced opinions and context-specific sentiments, which traditional methods may overlook.

Another area for improvement of real-time analytics**.** Currently, the system analyzes reviews after they are submitted, but implementing real-time processing could help administrators quickly identify emerging trends, potential issues, or suspicious activities. This feature would beneficial for companies aiming to provide fast responses customer feedback and enhance their decision-making process.

Additionally, the system could be expanded to include multi-platform review aggregation. This cross-platform integration would improve sentiment scoring and allow for a broader satisfaction.

Future may also involve personalized recommendation systems that suggest improvements based on analyzed reviews. By applying natural language processing and pattern recognition, the system could provide actionable insights to businesses or users, thereby making the feedback process more proactive rather than reactive.

Finally, there is scope for implementing robust security and privacy features. Safeguarding confidential user information and adhering to data protection laws enhances user confidence and system dependability. Incorporating features such as encrypted storage, secure authentication, and anonymized data analysis secure for widespread adoption.

REFERENCES

[1]. A. K. Vyas, A Comparative Analysis of Natural Language Processing Techniques for Sentiment Analysis, Int. J. Intell. Sys. Appl. Eng., vol. 12, no. 19s, pp. 1– , 2025
[2]. S. Baskar, A Comprehensive Review on Techniques in Sentiment Analysis for Improving Teaching and Learning through Students' Feedback, Next-Gen Comput. Syst. & Technol., vol. 1, no. 1, pp. 11–17, Oct. 2025.
[3]. Exploring the Effectiveness of ML and DL Algorithms for Sentiment Analysis: A Systematic Literature Review, Computers, Materials & Continua, vol. 84, no. 3, pp. 4105–4153, Jul. 2025.
[4]. C. Dev and A. Ganguly, "Sentiment Analysis of Review Data: A Deep Learning Approach Using User-Generated Content," Asian J. Electr. Sci., vol. 12, no. 2, pp. 28–36, Nov. 2023.
[5]. A review of Chinese sentiment analysis: subjects, methods, and trends, Artif. Intell. Rev., vol. 58, art. 75, Jan. 2025.
[6]. P. S. Ghatora, S. E. Hosseini, S. Pervez, M. J. Iqbal, and N. Shaukat, "Sentiment Analysis Reviews Using ML and Pre-Trained LLM," Big Data Cogn. Comput., vol. 8, no. 12, art. 199, 2024.
[7]. M. Shoqul Islam et al., "Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach," Artif. Intell. Rev., vol. 57, art. 62, Mar. 2024.
[8]. An analytical assessment of sentiment analysis trends and methods through systematic review and topic modeling, Decision Analytics J., vol. 17, Dec. 2025.
[9]. S. Rokhva, M. Alizadeh, and M. Abdollahi Shamami, Enhanced Sentiment Interpretation via a Lexicon-Fuzzy-Transformer Framework, arXiv:2510.15843, Oct. 2025.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr680

[10]. Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis," arXiv, Mar. 2024.
[11]. S. B. Linge, V. U. Sathe, P. S. Patil, and S. More, A Comprehensive Review of Current Methods, Progress, and Challenges in Sentiment Analysis, J. Interdiscip. Knowl., vol. 8, 2025,
[12]. Generalizing sentiment analysis: a review of progress, challenges, and emerging directions, Soc. Netw. Anal. Mining, 2025.
[13]. M. H. Chauhan and D. D. Vyas, Advancements in Sentiment Analysis – A Comprehensive Review of Recent Techniques and Challenges, The Scientific Temper, vol. 16, no. Spl-1, May 2025.
[14]. N. Sahu and K. Sharma, Sentiment Analysis Using Machine Learning, Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 11, no. 5, pp. 133–137, Oct. 2025.
[15]. R. A. García-Hernández et al., A Systematic Literature Review of Modalities, Trends, and Limitations in Emotion Recognition, Affective Computing, and Sentiment Analysis, Appl. Sci., vol. 14, no. 16, 7165, 2024.
[16]. Y. Mao, Q. Liu, and Y. Zhang, Sentiment Analysis Methods, Applications, and Challenges: A Systematic Literature Review, J. King Saud Univ. – Comput. Inf. Sci., vol. 36, no. 4, 102048, Apr. 2024.