International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 25 Issue 01, JAN, 2025

LEVERAGING PRODUCT CHARACTERISTICS FOR ONLINE COLLUSIVE DETECTION IN BIG DATA TRANSACTIONS

Mr.K.V.Rajesh[1],Gorla Rishitha[2], Kurva Shivani[3], Kansampally Shireesha[4]

[1]Associate Professor, School of CSE ,Malla Reddy Engineering College For Women(Autonomous Institution), Maisammaguda, Dhulapally,Secunderabad,Telangana-500100

[234]UG Student, Department of CSE,Malla Reddy Engineering College for Women, (Autonomous Institution), Maisammaguda,Dhulapally,Secunderabad,Telangana-500100

Mail Id: kvrajeshh@gmail.com

ABSTRACT

In the rapid growth of e-commerce, online fraud especially in terms of using false usernames has become one of the major concerns both to the platforms and to users. Usually, fraudsters get confidence with the username system which also gets customers and leads them to a loss and the subsequent mistrust of the platform. The paper presents a conceptual framework for detecting fraud through a set of individual and transaction-related signals analysis. The framework introduces two features of the product: product type and product appearance, to improve the accuracy of fraud detection. This framework uses data mining techniques to extract and analyze key features from real-world datasets, aiming to distinguish legitimate transactions from fraudulent ones. The experimental results show that the proposed signals can effectively identify fraud, thereby providing a powerful means to enhance the security and integrity of online shopping. The study highlights the role of big data technology in improving decision making and building a proper reputation system, thus creating a safer e-commerce environment.

Keywords : E-commerce, online fraud, fraud detection, reputation system, data mining, big data, transaction indicators, product features, machine learning, fraud prevention, data analysis.

I.INTRODUCTION

The rapid advancement of the Internet of Things (IoT) has led to the generation of vast amounts of data, offering tremendous opportunities for various industries, particularly in online commerce. Big data technologies, such as data mining and machine learning, enable e-commerce platforms to harness this data to gain valuable insights, enhance decision-making, and optimize business operations. Through IoT and big data, online platforms can improve the transaction process, create a healthy shopping environment, and, in the end, enhance sales.The high efficiency and low cost of IoT have boosted the growth of online shopping. In 2018, the CNNIC published the 2017 China Internet Development Statistic Report, showing that there were more than 772 million online users in China alone as of 2017. Despite this growth, fraud by fake sellers exploiting the reputation systems on online platforms like Taobao still exists. Since there is millions of fraudulent transactions done daily, online platforms stand risks not only financially but also towards a loss of customer satisfaction. Big data really matters in the war against fraudulent behavior because it supplies with tools that would ensure them to detect fraud behavior patterns.One of the biggest problems with e-commerce is identifying participants, as these transactions are virtual. Thus, a consumer cannot see a product before purchasing it or even evaluate its quality. Online buyers have to rely on less and often incomplete information than that which is available to sellers. Most of the systems implement

reputation systems, by which buyers can rate their sellers based on historical transaction ratings. These reputation systems are of utmost importance to creating trust among buyers and sellers but vulnerable to manipulation.Collusion fraudsters inflate reputation scores in order to be considered more reliable than they really are. This is how fraudulent behavior can break down reputation systems and hurt the consumers. Thus, e-commerce needs big data for identifying and preventing frauds. Fraudsters are motivated by the economic benefits associated with high reputation scores, which can drive higher sales and profits. However, the lack of effective monitoring and the difficulty of detecting collusion make it challenging to prevent fraud.While reputation systems are a valuable tool for identifying reliable sellers, they are vulnerable to manipulation. Detecting and preventing fraudulent transactions has become a pressing task for e-commerce platforms. New type product nature is suggested as new features while producing detection of fraudulent behavior for that. Combining that and combining other user and transaction related characteristics, we offer to improve the accuracy with such a model. The contributions of this paper include introducing new fraud detection features, validating them using a real-world dataset, and giving policy recommendations for online platforms about how to ensure the integrity of their reputation systems.

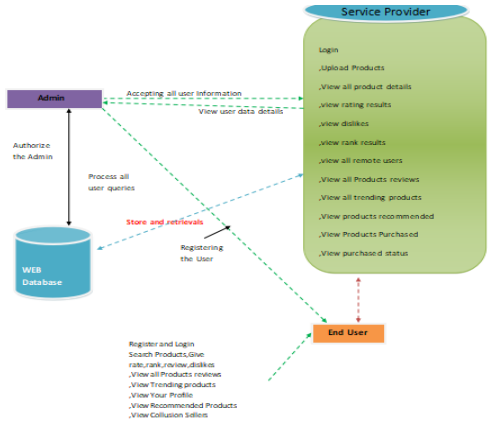

Fig 1 : System Architecture

II.RELATED WORK

Detecting financial fraud using data mining techniques: a decade review from 2004 to 2015
Author: M. Albashrawi (2016)
This paper provides a comprehensive review of financial fraud detection techniques using data mining, covering the period from 2004 to 2015. It evaluates various methods, including classification, clustering, and anomaly detection, highlighting their applications in detecting financial fraud and the evolution of these techniques over the years.

A risk-based ranking of product listings at online auction sites for non-delivery fraud prediction
Author: V. Almendra (2013)
This research focuses on online auction fraud, specifically non-delivery fraud. The author proposes a risk-based ranking system that uses product listings and buyer behaviors to predict fraudulent activities. The study highlights how auction platforms can identify suspicious listings and prioritize them for further review.

Fraud Detection by Human Agents: A Pilot Study
Author:V. Almendra & D. Schwabe (2009)
This paper investigates fraud detection by human agents in e-commerce environments. It explores how human experts can identify fraudulent transactions that automated systems might miss. The study emphasizes the importance of combining human expertise with automated fraud detection tools for improved accuracy.

Spotting fake reviewer groups in consumer reviews
Author:A. Mukherjee, B. Liu, & N. Glance (2012)
The authors address the problem of fake reviews in online platforms. They propose methods to identify groups of fake reviewers who engage in coordinated efforts to distort the reputation of products and services. The paper employs techniques such as social network analysis to detect fraudulent review patterns.

Graph Theory and Combinatorics
Author: B. Bollobás (1984)

Although this reference is not directly related to fraud detection, it provides fundamental insights into graph theory and combinatorics, which are applicable to fraud detection in network-based systems. The concepts in this book can be used to analyze patterns and relationships in data, such as identifying suspicious behavior in financial transactions or online networks.

Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior
Author(s): S. Ba & P. Pavlou (2002)
This paper examines how trust-building technologies, such as reviews, ratings, and guarantees, influence buyer behavior in electronic markets. The authors show that these trust mechanisms reduce fraud by creating a more reliable and transparent marketplace, thus affecting price premiums and overall buyer decisions.

III.IMPLEMENTATION

The implementation of a fraud detection framework for an e-commerce platform using data mining and machine learning techniques involves several key steps. It begins with data collection, including a comprehensive dataset of legitimate and fraudulent transactions. Such a dataset would include user attributes, such as the username, account number, and transaction frequency; transaction attributes, like product details and the number of transactions; and behavioral attributes, such as transaction patterns. After gathering data, all data collected would require processing, such as data cleaning handling missing and duplicate values), matching numeric features, and encoding for segment variables. Machine learning is critical to any fraud-detection technique: this encompasses feature extraction into relevant factors, like type of products, description of products as extracted by natural language processing the description of the products, analysis of name tags for indication of any fraudulence.

Other behavioral indicators, such as interaction patterns and transaction frequency, should be considered. Then, once the features are consistent, choose and train a model using the machine learning techniques. It can be achieved by selecting supervised models such as Random Forest, XGBoost, or Logistic Regression to classify transactions as fraud or valid, and then unsupervised models like Isolation Forest or k-Means to identify outliers and anomalies. Once the model is trained, it is necessary to validate its performance by using some metrics such as precision, accuracy, recall, and F1 score. It should be cross-validated so that the model will be robust. Once this model is refined, it can be deployed into production and suspicious transactions can be flagged in no time. Continuous monitoring and periodic retraining of models are essential to adapt to new fraud patterns.

In addition, e-commerce platforms can be advised to strengthen their reputation systems to prevent manipulation and implement fraud detection methods. In short, the introduction of big data technologies and machine learning algorithms can improve fraud detection in e-commerce, thereby creating a safer environment for consumers and businesses.

IV.ALGORITHMS

The fraud detection algorithms of e-commerce transactions are systematic. They use data mining and machine learning techniques to identify fraudulent activities. The algorithm starts by collecting data from the e-commerce site, which includes user attributes (user ID, account name, account number), product attributes (product ID, description, price), transaction attributes (transaction ID), customer ID and customer, amount, and behavioral attributes (transaction frequency, account activity). The data is pre-processed, including cleaning (management of missing and duplicated values), numerical data matching, and identification of key features: product type and customer ID. After this pre-processing stage, the relevant features were extracted. These include details about the product, the user's behavior, and his or her transaction patterns. Build in additional features, including location information (for example, time between purchases), and behavioral cues (for example, sudden changes in name tags) to enhance the

precision of the model. Data set is split into training set and test set. These algorithms employ models like random forest, logistic regression, and gradient boosting for supervised learning, while anomaly detection and unsupervised learning use models like sparse forests. The model is then trained on the training data and optimized with hyperparameter tuning techniques such as grid search. After training, the performance of the model is evaluated using metrics such as precision, accuracy, recall, F1 score, and confusion matrix. The model feeds into operations of training in real time to classify the incoming transaction as either fraudulent or legitimate for predefined thresholds, for example; Suspicious transactions are automatically flagged, and the administrators receive alerts to check further. The effectiveness of the model over time depends on this monitoring and retraining with newer data. This system can, therefore, make suggestions for the name system's improvement to avoid fraud manipulation and thereby enhance fraud prevention measures on that platform.

V. RESULTS

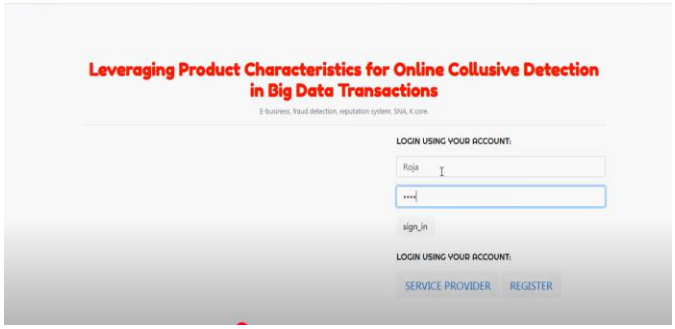

Fig 1: Login Page

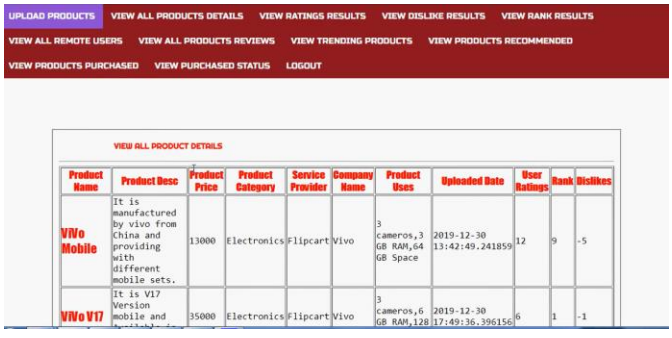

Fig 2: Product Details

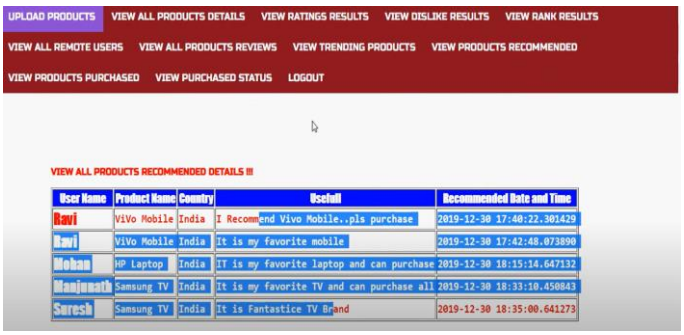

Fig 3: Recommended Details

Fig 3:line chart

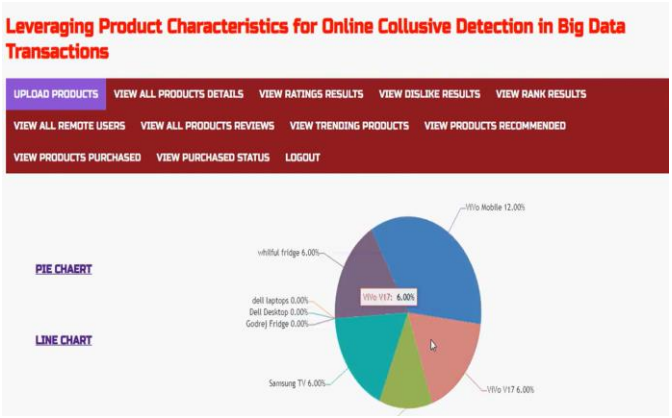

Fig 4:Pie Chart

Fig 5:Accuracy Bar Chart

VI.CONCLUSION

This work provides a framework for detecting fraud on an e-commerce platform. Three aspects play an important role. Firstly, this work presents two new features which can detect fraudulent transactions; it further combines those with other features related to the user, making the detection of fraud highly accurate. It may then be generalized for detection across various platforms of different fraud types. We validate the usefulness of the model with real-world data sets, which demonstrates its effectiveness in fraud identification and valuable insights for online e-commerce platforms. Thirdly, we present some recommendations for policy that will help ensure online reputation systems are secure and describe their importance in reducing the fraud rate and building consumer trust. Our findings show remarkable behavioral differences between legitimate and fraudulent users, particularly the ways that fraudsters exploit popular platforms by posting high ratings and detailed reviews despite conducting low-value transactions. Moving forward, future research should concentrate on the development of more general fraud detection models that can be applied to a variety of e-commerce platforms. In addition, data from multiple platforms could be used to assess the financial impact of fraud and refine detection models. Since fraudsters continue to evolve their strategy, there is a continuous need to build adaptive models capable of identifying new and emerging frauds.

REFERENCES

[1] Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: a decade review from 2004 to 2015.

[2] Almendra, V. (2013). Finding the needle: A risk-based ranking of product listings at online auction sites for non-delivery fraud prediction. Expert Systems with Applications, 40(12), 4805–4811.

[3] Almendra, V., & Schwabe, D. (2009). Fraud Detection by Human Agents: A Pilot Study. E-Commerce and Web Technologies. Springer Berlin Heidelberg. 10th International Conference, Linz, Austria, September 1–4, Springer, New York, NY, 2009, 300–311.

[4] APAMukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. International Conference on World Wide Web (pp.191-200). ACM.

[5] B. Bollob´as,(1984). in Graph Theory and Combinatorics: Proc. Cambridge Combinatorial Conf. in honour of Paul Erd˝os. B. Bollob´as, ed. Academic Press, NY,pp. 35-37.

[6] Ba, S., and Pavlou, P. (2002). Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. Mis Quarterly, 26(3), 243-268.

[7] Becker, G. S. (1968). Crime and punishment: an economic approach. Journal of Political Economy, 76(Volume 76, Number 2), 169-217.

[8] Berlusconi, G. (2017). Social Network Analysis and Crime Prevention. Crime Prevention in the 21st Century. Springer International Publishing.

[9] Blume, M., Weinhardt, C., and Seese, D. Using network analysis for fraud detection in electronic markets. In T. Dreier, R. Studer, and C. Weinhardt (eds.). Information Management and Market Engineering, Volume 4 of Studies on eOrganisation and Market Engineering, vol. 4, University Atsverlag Karlsruhe, Germany, 2006, 101–112.

[10]Bolton, G. E., Katok, E., & Ockenfels, A. (2005). How effective are online reputation mechanisms? an experimental study. Management Science, 50(3).

[11]Borgatti, S. P., Jones, C., & Everett, M. G. (2004). Network Measures of Social Capital. (Vol.21).

[12]Bradley, P. S. Mangasarian, O. L., and Street, W. N. (1997). Clustering via concave minimization. Advances in Neural Information Processing Systems -9, 368--374.

[13]Carrington PJ (2011) Crime and social network analysis. In: Scott JP, Carrington PJ (eds) The SAGE handbook of social network analysis. SAGE Publications, London, pp 236–255

[14]Chae, M., Shim, S., Cho, H., & Lee, B. (2007). An empirical analysis of fraud detection in online auctions: Credit card phantom transaction. In Proceedings of the 40th annual Hawaii international conference on system sciences, HICSS ’07 (pp. 155a). IEEE Computer Society.

[15]Chang, W.-H., & Chang, J.-S. (2010). Using clustering techniques to analyze fraudulent behavior changes in online auctions. In International conference on networking and information technology (ICNIT) (pp. 34–38).

[16]Chang, W.-H., & Chang, J.-S. (2011). A novel two-stage phased modeling framework for early fraud detection in online auctions. Expert Systems with Applications, 38(9), 11244–11260.

[17]Chau, D. H., Pandit, S., & Faloutsos, C. (2006). Detecting Fraudulent Personalities in Networks of Online Auctioneers. European 2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission.