A Comprehensive Review on Effective Fraud Detection in E-Commerce Using Machine Learning Techniques and Big Data Analytics

[1]Laxman Kumar Mahto, [2]Aashish Kumar Tiwari, [3]Dr. Saurabh Mandloi

MTech Scholar, Department of Computer science and Technology, Sam College of Engineering & Technology, Bhopal

Asst. Prof., Department of Computer science and Technology, Sam College of Engineering & Technology, Bhopal

Asst. Prof., Department of Computer science and Technology, Sam College of Engineering & Technology, Bhopal

1laxman91mahto@gmail.com , [2]aashish.tiwari7898@gmail.com , [3]saurabhm.research@gmail.com

Abstract

The rapid expansion of e-commerce has revolutionized global trade, offering unprecedented convenience and accessibility. However, this growth has simultaneously fueled a rise in sophisticated fraud activities, posing significant financial, operational, and security challenges for online businesses. Traditional fraud detection methods, primarily rule-based systems, have become increasingly inadequate due to their limited adaptability, high false-positive rates, and inability to process large volumes of dynamic data. Recent advancements in machine learning (ML) and big data analytics offer promising solutions capable of identifying complex fraud patterns, adapting to evolving attack strategies, and enabling real-time decision-making. This review paper provides a comprehensive examination of the role of ML and big data technologies in e-commerce fraud detection. It explores the characteristics of fraud data, discusses widely adopted ML approaches—including supervised, unsupervised, and deep learning models—and evaluates the strengths and limitations of each. The paper also analyzes key big data frameworks, such as Hadoop, Spark, Flink, and Kafka, that support scalable and low-latency fraud detection architectures. Additionally, major challenges including data imbalance, concept drift, privacy concerns, scalability limitations, and adversarial threats are critically examined. By synthesizing findings from recent research and industrial practices, this review highlights current gaps and proposes future directions such as federated learning, explainable AI, and adversarially robust models. The insights presented aim to guide researchers and practitioners in developing efficient, intelligent, and resilient fraud detection systems for the evolving e-commerce landscape.

Keywords: Big data, Fraud detection, Machine Learning, Fraud Types, Challenges in Fraud Detection.

I. Introduction

The exponential growth of online transactions has led to the generation of massive amounts of data, necessitating sophisticated cyber-infrastructure and information technology methods for effective exploitation and analysis. However, this digital expansion has simultaneously made both individuals and businesses vulnerable to financial fraud, a pervasive global problem. Financial fraud is defined as the act of obtaining financial gains through illegal and fraudulent means. It can be perpetrated across various financial sectors, including banking, insurance, corporate, and taxation [1].

In recent times, organizations have faced a growing challenge from various financial crimes, such as money laundering and fraudulent financial transactions. Despite numerous efforts to curb fraud, the issue persists, severely impacting the economy and society daily, resulting in substantial monetary losses. Historically, many fraud identification techniques were proposed, but the bulk of older processes were manual, proving inefficient, costly, inaccurate, and time-consuming [2].

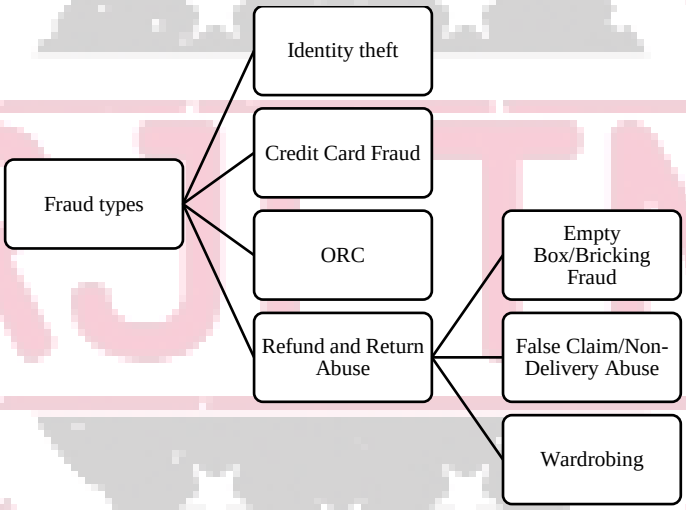
The battle against fraud has been significantly aided by advancements in artificial intelligence (AI) and machine learning (ML). These technologies are supporting fast digitization and revolutionizing fraud prevention efforts. ML and AI algorithms empower companies to sift through vast amounts of data to find patterns and anomalies that may suggest fraudulent activity. These technologies are crucial as they enable advanced data analytics,

anomaly detection, and predictive modeling. Leveraging these tools allows organizations to proactively identify and mitigate fraud risks, thereby safeguarding their operations and stakeholders [3]. The intersection of ML and big data technologies offers a promising frontier for developing robust and adaptive fraud detection systems in e-commerce. However, despite significant advancements, several critical challenges persist, including data imbalance, concept drift, privacy concerns, and the need for transparent and explainable detection models. These challenges continue to motivate researchers and industry practitioners to explore innovative, scalable, and ethical solutions [4, 5].

This review paper aims to provide a comprehensive analysis of current approaches, technological developments, and research trends in e-commerce fraud detection. It examines the evolution from traditional fraud prevention methods to sophisticated ML-driven and big data-enabled techniques, assesses the strengths and limitations of existing models, and highlights key research gaps and future directions. By synthesizing insights from both academia and industry, this paper seeks to contribute to the ongoing development of effective, reliable, and intelligent fraud detection systems within the e-commerce ecosystem.

II. Domains in Fraud detection

The landscape of e-commerce and retail fraud is constantly evolving, requiring sophisticated, adaptive measures. The three most prevalent and challenging forms include identity theft, credit card fraud, and Organized Retail Crime (ORC).



Different Frauds in the market

- Identity Theft (Account Takeover - ATO): This type of fraud involves a malicious actor gaining unauthorized access to a legitimate customer's account to make purchases, change shipping addresses, or harvest sensitive data. The sophistication of this method lies in how fraudsters mimic legitimate behavior. Conventional detection methods, which rely on simple checks like IP address comparison or single velocity rules, are easily bypassed [6]. For instance, a rule that flags all purchases made from a new country might be too restrictive. Advanced Detection Need: Machine Learning (ML) algorithms are essential here. They can analyze vast amounts of data to establish a baseline of normal user behavior—including typical browsing speed, scrolling patterns, time of day for transactions, and item preferences. Irregularities, such as an abrupt change in purchase amount or the speed at which a customer navigates the checkout process (an example of behavioral biometrics), become statistically significant deviations that ML models can spot in real-time [7].
- Credit Card Fraud (Payment Fraud): This involves the unauthorized use of stolen or compromised credit card details (e.g., card number, CVV, expiration date) to make illegal online purchases. This is often executed via card-not-present (CNP) transactions. Traditional rule-based algorithms rely on static rules, such as flagging transactions exceeding a certain dollar limit or those originating from a high-risk country list. However, fraudsters quickly adapt by making small, frequent purchases (card testing) or using new proxy servers to bypass these simple checks [8]. Advanced Detection Need: ML models excel at reading complex, non-linear patterns across multiple features simultaneously a capability beyond simple

threshold rules. They can detect subtle correlations, like a specific combination of low-value items, a certain delivery address structure, and a particular browser type that collectively indicates a high probability of fraud [9].

- Organized Retail Crime (ORC): Unlike opportunistic, individual fraud, ORC involves professional, criminal groups conducting theft and fraud for commercial gain. While traditionally focused on physical store theft, ORC has broadened its scope to e-commerce. This includes large-scale return fraud, "buy online, pick up in store" (BOPIS) fraud, and triangulation fraud, where fraudsters act as legitimate sellers using stolen cards to fulfill orders [10]. The complexity requires a holistic fraud-management framework, which is the focus of advanced e-commerce security architectures.
- Refund and Return Abuse: This category involves exploiting lenient return and refund policies for financial gain without involving stolen payment details. It is considered a form of first-party fraud. Wardrobing is a fraud type in which the customer purchases an item, uses it (especially clothing for an event), and then returns it for a full refund, claiming it was never used or that they simply changed their mind [11]. Empty Box/Bricking Fraud is another type The customer returns an empty box, a box filled with trash, or an old/broken version of the product (e.g., returning an electronic device after stripping it of valuable internal components, known as "bricking") while claiming a refund for the new, high-value item [12]. False Claim/Non-Delivery Abuse in which the customer falsely claims an item was never delivered or arrived damaged/not as described, particularly for low-value items where merchants may issue a refund without requiring a return to save on logistics costs. Detecting this abuse requires analyzing non-payment data, such as the customer's return history frequency, the difference between purchase and return locations, the return reason codes, and the total lifetime value versus the lifetime return value for a specific user ID to flag chronic policy abusers [13,14].

III. Challenges in ML and Big Data for Fraud Detection

E-commerce fraud datasets are typically highly imbalanced, with fraudulent transactions representing a very small fraction of the total data. In many real-world scenarios, less than 0.5% of transactions are labeled as fraudulent, making it extremely difficult for machine learning models to learn meaningful fraud patterns [15]. Standard classifiers often become biased toward predicting the majority “legitimate” class while failing to identify rare fraudulent events. Techniques such as oversampling, undersampling, and SMOTE may help, but they introduce risks like overfitting or distortion of data distributions. Advanced methods such as cost-sensitive learning and anomaly detection models offer improvements, yet tackling extreme imbalance remains one of the biggest obstacles in operational fraud detection systems [16].

Fraudsters continuously modify their tactics to evade detection, leading to what is known as *concept drift*, where the underlying distribution of legitimate and fraudulent behavior changes over time [17]. Models trained on historical data often degrade in performance as new fraud patterns emerge. Static models quickly become obsolete, requiring continuous retraining, adaptive learning methods, and streaming analytics to maintain effectiveness. Concept drift poses a severe challenge because fraud evolves faster than the capability of many organizations to update and deploy new fraud detection models [18].

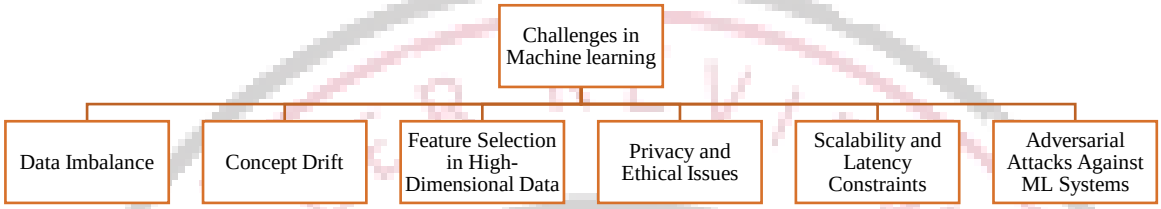
Fraud detection systems rely on diverse features such as device information, user behavior, transaction metadata, network relationships, and temporal sequences. This results in high-dimensional datasets that increase computational complexity and introduce noise, making it difficult to isolate the most relevant predictors [19]. Feature engineering becomes an intensive task that demands domain expertise and computational resources. Methods such as PCA, autoencoders, and embedded feature-selection techniques help reduce dimensionality, but extracting meaningful and stable features for evolving fraud patterns remains a persistent challenge [20].

To detect fraud effectively, models often require sensitive personal or financial data, raising concerns regarding consumer privacy, regulatory compliance, and ethical data usage. Regulations such as GDPR and CCPA restrict how data can be collected, stored, or shared across systems and jurisdictions [21]. Additionally, ML-based fraud detection systems risk embedding biases from historical data, potentially leading to unfair or discriminatory outcomes. Ensuring transparency, explainability, fairness, and responsible handling of user data is essential yet complex in high-scale fraud detection environments [22].

Fraud detection systems must operate in real time to prevent financial loss before transactions are completed. Processing billions of transactions, logs, and user interactions requires high-throughput big-data architectures capable of handling massive scale with minimal latency [23]. Traditional ML models often fail to meet real-time

decision-making constraints when deployed in distributed environments. Big data frameworks like Apache Kafka, Spark Streaming, and Flink support large-scale streaming analytics, but integrating ML models into these pipelines in a low-latency manner remains technically challenging [24].

Fraudsters increasingly employ adversarial techniques designed to deceive machine learning models. By manipulating input features—such as modifying transaction amounts, masking device information, or using automated bots—they exploit model weaknesses to avoid detection [25]. More sophisticated adversarial attacks involve probing model decision boundaries or reverse-engineering deployed fraud detection systems. Defensive strategies include adversarial training, model hardening, ensemble defenses, and monitoring abnormal patterns, but adversarial resilience is still a developing field in fraud analytics [26].



Figures2: Challenges in using ML for Fraud detection

Table 1: Comparative Assesment of different Approaches used in fraud detection

Ref.	Challenge Focus	Problem / Dataset	Methods / Approach	Key Findings
[27]	Data imbalance, deep learning	Financial fraud datasets with severe class imbalance in real institutions	Surveys and evaluates deep learning architectures under imbalanced conditions; discusses loss functions and sampling strategies.	Shows that standard deep models overfit majority classes; recommends specialized loss functions and hybrid resampling to improve minority (fraud) detection.
[28]	Data imbalance, generative models	Imbalanced credit-card fraud data	Proposes generative models to create synthetic fraud samples and compares several generators.	Demonstrates that generative oversampling significantly improves recall and AUC for rare fraud cases vs. classical resampling.
[29]	Data imbalance, preprocessing	Multiple real and benchmark imbalanced datasets	Systematic comparison of preprocessing pipelines (scaling, sampling, feature selection) for ML and DL models.	Finds that combining normalization with targeted oversampling yields consistent gains and reduces false negatives in fraud-like settings.
[30]	Concept drift in streaming fraud	Streaming credit/debit transaction data	XGBoost-based fraud detector with drift-aware updates for streaming environments.	Shows that periodically retrained models with drift handling outperform static models on evolving fraud streams.
[31]	Concept drift, scalability	Multiple streaming fraud datasets (Adaptive Random Forest)	Two-stage framework: offline initialization + online drift detection (DDM, EDDM, ADWIN) with incremental training.	Achieves high AUC while reducing retraining frequency; ADWIN + Adaptive RF provides best trade-off between speed and accuracy under drift.
[32]	Streaming, scalability & latency	Streaming credit-card transactions	Clusters cardholders and uses sliding windows to analyze behaviour for real-time detection.	Highlights practical pipeline design for near real-time fraud detection and shows improved detection on large streaming datasets.

Recent research has increasingly focused on enhancing fraud detection systems through privacy-preserving machine learning and adversarially robust modeling, addressing gaps in data sharing, regulatory constraints, and system vulnerability. Kanamori et al. [33] proposed a federated learning framework, DeepProtect, to address the limitations of centralized fraud detection in the financial sector. Their approach enables multiple banks to collaboratively train models without exchanging raw customer data, integrating secure multiparty computation and differential privacy to protect sensitive information. Evaluations conducted on real datasets from Japanese banks demonstrated that federated models can achieve performance levels comparable to centralized systems while fully maintaining privacy compliance. Building on this foundation, Emmanuel et al. [34] extended federated learning to broader ecosystems involving e-commerce merchants, payment gateways, and financial institutions. Their work emphasizes the value of collaborative modelling in detecting cross-merchant fraud rings and push-payment scams, challenges that individual entities cannot effectively capture in isolation. By addressing issues such as heterogeneous data distributions and communication overhead, they showed that federated approaches significantly enhance detection accuracy across diverse platforms.

In contrast, Kuleshov et al. [35] examined the rising threat of adversarial attacks targeting fraud detection models deployed in real-world environments. Their findings reveal that fraudsters can exploit weaknesses in model decision boundaries through carefully crafted inputs and through probing operational constraints, leading to significant misclassification rates. They propose domain-specific adversarial threat models and highlight the need for adaptive, robust defences. Complementing this, Cartella et al. [36] investigated adversarial vulnerabilities in tabular models commonly used for fraud detection, such as tree-based classifiers and neural networks. Their results demonstrate that even small, structured perturbations to transactional or behavioural features can bypass detection systems, while existing defenses offer limited resilience. Collectively, these studies emphasize the necessity of balancing privacy, collaboration, and robustness when designing next-generation fraud detection frameworks.

IV. Big Data Analytics in E-Commerce Fraud Detection

Big data analytics plays a central role in modern e-commerce fraud detection due to the scale, speed, and complexity of transactional data generated across digital platforms. Fraud-related datasets are inherently characterized by high dimensionality, as each transaction can contain hundreds of attributes, including device identifiers, customer behavior metrics, geolocation points, timestamps, and network-based relational features. Such high-dimensional data requires advanced preprocessing, dimensionality reduction, and feature-selection techniques to maintain model efficiency and interpretability [37]. Another critical characteristic is extreme class imbalance**,** where fraudulent transactions typically account for a very small minority of all data—often below 0.5%. This imbalance challenges traditional machine learning algorithms, which tend to favor majority-class patterns, thereby missing subtle fraud signals. Techniques such as anomaly detection, cost-sensitive learning, and synthetic oversampling have been shown to improve fraud identification under severe imbalance [38]. Additionally, e-commerce systems require real-time processing, as businesses must detect and block fraudulent activities within milliseconds to avoid monetary loss and preserve customer trust. This real-time need demands highly optimized, low-latency architectures capable of continuous monitoring under massive data loads [39].

To meet these challenges, organizations leverage a variety of big data frameworks and tools. The Hadoop ecosystem provides scalable storage and distributed batch processing through HDFS and MapReduce, enabling efficient analysis of historical fraud data and large-scale training of ML models [40]. Building on this foundation, Apache Spark MLlib offers fast, in-memory processing for iterative machine learning tasks, making it well-suited for fraud classification, clustering, and anomaly detection at scale. For real-time fraud detection, Apache Kafka acts as a high-throughput message broker for ingesting transactional streams, while Apache Flink processes these streams with low latency and high fault tolerance, supporting immediate anomaly scoring and event-based alerting [41]. Moreover, NoSQL databases such as MongoDB and Cassandra provide flexible schemas and fast read/write operations, enabling efficient handling of semi-structured and unstructured fraud data used in both historical and real-time analyses [42].

These capabilities support the design of real-time fraud detection architectures, which integrate streaming ingestion, distributed processing, and machine learning inference into continuous analytics pipelines. In such setups, Kafka or similar systems collect incoming transactions, which are then evaluated through Flink or Spark Streaming engines using pre-trained models. A major architectural consideration is balancing batch versus real-time processing. Batch pipelines analyze large historical datasets to update fraud models periodically, while real-time layers deploy these models to score each transaction instantly. Most modern systems adopt hybrid architectures**,** combining both approaches to ensure accuracy, adaptability, and minimal latency [43]. These hybrid

designs enable dynamic responses to evolving fraud patterns, leveraging batch insights for robust learning while maintaining always-on real-time protection [44][45].

V. Conclusion

E-commerce fraud continues to evolve in complexity, making effective detection a critical priority for online businesses and financial institutions. This review has demonstrated that machine learning and big data analytics represent powerful tools for addressing the limitations of traditional rule-based fraud detection systems. By leveraging high-dimensional transactional data, behavioral metrics, and real-time processing capabilities, ML-driven approaches can uncover hidden patterns, respond dynamically to emerging threats, and significantly improve fraud detection accuracy. Big data frameworks such as Hadoop, Spark, Flink, and Kafka further enhance scalability and operational efficiency, enabling the deployment of real-time fraud detection pipelines capable of handling millions of events per second.

Despite these advances, several challenges remain. Data imbalance continues to hinder model performance, concept drift requires continuous learning and adaptation, and privacy regulations restrict data sharing across organizations. Additionally, adversarial attacks on ML models expose critical vulnerabilities that must be addressed through robust training and advanced defense strategies. Future research must focus on developing explainable and transparent ML models, privacy-preserving approaches such as federated learning, and hybrid architectures that combine batch and streaming analytics for optimal performance. Strengthening adversarial resilience and ensuring ethical and fair model behavior will also be essential. Overall, the integration of ML and big data analytics offers tremendous potential for building intelligent, scalable, and secure fraud detection systems. With continued innovation and responsible deployment, these technologies can significantly reduce fraud risk and contribute to a safer and more trustworthy e-commerce ecosystem.

References

[1] Kshetri, N. (2021). *The economics of e-commerce frauds*. In Big Data's Big Potential in Developing Economies. Palgrave Macmillan.

[2] Statista. (2023). *E-commerce losses to online payment fraud worldwide*. https://www.statista.com

[3] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review. *Decision Support Systems, 50*(3), 559–569.

[4] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176.

[5] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98–115.

[6] Carcillo, F., Dal Pozzolo, A., Le Borgne, Y. A., Caelen, O., Mazzer, Y., & Bontempi, G. (2019). SCARFF: A scalable framework for streaming fraud detection with Spark. *Information Fusion, 41*, 182–194.

[7] N. K. Singh and V. D. Soni, "A review of machine learning techniques for credit card fraud detection," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 5, pp. 110-116, 2018.

[8] V. S. Ramachandran, S. B. Subrahmanyam, and D. R. N. D. R. Reddy, "Machine learning techniques for behavioral biometrics in fraud detection," in *2020 IEEE Int. Conf. on Commun. and Signal Process. (ICCSP)*, Chennai, India, 2020, pp. 297-301.

[9] A. K. Jha and V. K. Jha, "Challenges and approaches for credit card fraud detection in E-commerce," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 12, pp. 3173-3179, 2019.

[10] Y. P. W. Gunawardena and P. C. H. Fernando, "Real-time credit card fraud detection using machine learning and deep learning," in *2021 Int. Res. Conf. on Smart Comput. and Systems (ICSCS)*, Colombo, Sri Lanka, 2021, pp. 1-6.

[11] National Retail Federation, "Organized Retail Crime," 2024. [Online]. Available: https://nrf.com/topics/retail-security/organized-retail-crime (Reference concept, specific URL and access date would be used in a final paper).

[12] J. K. J. K. John and M. M. M. S. Sreeja, "A comparative study on different machine learning algorithms for credit card fraud detection," in *2021 Int. Conf. on Adv. Comput. and Commun. Systems (ICACCS)*, Coimbatore, India, 2021, pp. 917-922.

[13] F. L. Leite, J. M. S. Loures, and M. R. M. F. de Oliveira, "Machine learning approach for false decline reduction in payment card systems," in *2022 IEEE Int. Conf. on Ind. Eng. and Eng. Manag. (IEEM)*, Malaysia, 2022, pp. 1297-1301.

[14] G. K. Sahoo, B. P. Acharya, and S. Varma, "A survey on real-time fraud detection systems," in *2022 Int. Conf. on Adv. Comput. and Comm. (ICACC)*, Bhubaneswar, India, 2022, pp. 1-6.

[15] Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8), 3784–3797.

[16] Haibo, H., Yang, B., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling for imbalanced learning. *IEEE IJCNN*, 1–8.

[17] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1–37.

[18] Baena-García, M., et al. (2006). Early drift detection method. *ECML/PKDD Workshop on Knowledge Discovery from Data Streams*, 77–86.

[19] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16–28.

[20] Hinton, G., & Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504–507.

[21] Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR)*. Springer.

[22] Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and Machine Learning*. MIT Press.

[23] Ramaswamy, S., et al. (2017). Streaming fraud detection in large-scale e-commerce systems. *ACM SoCC*, 111–124.

[24] Carbone, P., et al. (2015). Apache Flink™: Stream and batch processing in a single engine. *IEEE Data Engineering Bulletin*, 38(4), 28–38.

[25] Goodfellow, I., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *ICLR*.

[26] Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84, 317–331.

[27] Chen, Y., et al. (2025). Deep learning in financial fraud detection: Innovations and challenges with imbalanced datasets. *Journal of Finance and Data Science*.

[28] Tayebi, M., et al. (2025). Generative modeling for imbalanced credit card fraud detection. *AI*, 5(1), Article 9.

[29] Abu Elfetouh, A., et al. (2025). Enhancing fraud detection in imbalanced datasets using data preprocessing for machine and deep learning algorithms. *Egyptian Informatics Journal* (or equivalent source as per journal site).

[30] Shahapurkar, A. S., & Patil, R. (2023). Concept drift and machine learning model for detecting fraudulent transactions in streaming environment. *International Journal of Electrical and Computer Engineering*, *13*(5), 5560–5568.

[31] Yelleti, V. (2025). ROSFD: Robust online streaming fraud detection with resilience to concept drift in data streams. *arXiv preprint arXiv:2504.10229*.

[32] Dornadula, V. N., & Geetha, S. (2019). Credit card fraud detection using machine learning: A systematic survey. *Procedia Computer Science*, *165*, 631–641 (plus streaming-data method as detailed in the paper).

[33] Kanamori, S., et al. (2022). Privacy-preserving federated learning for detecting fraudulent financial transactions in Japanese banks. *Journal of Information Processing*, *30*, 789–803.

[34] Emmanuel, M., et al. (2025). Federated learning for privacy-preserving financial fraud detection in e-commerce and push-payment systems. *SSRN Working Paper*.

[35] Kuleshov, V., et al. (2023). Adversarial learning in real-world fraud detection. *arXiv preprint arXiv:2307.01390*.

[36] Cartella, F., et al. (2021). Adversarial attacks for tabular data: Application to fraud detection. In *Proceedings of the Italian Workshop on Machine Learning and Data Mining* (CEUR-WS Vol. 2808)

[37] Zhao, R., & Hryniewicki, M. (2018). Dimensionality reduction and machine learning methods for anomaly detection. *Applied Soft Computing*.
[38] Douzas, G., & Bacao, F. (2018). Effective data generation for imbalanced learning using GANs. *Expert Systems with Applications*.
[39] Sadgali, I., et al. (2019). Real-time big data processing for fraud detection in e-commerce. *Procedia Computer Science*.
[40] White, T. (2015). *Hadoop: The Definitive Guide*. O'Reilly Media.
[41] Carbone, P., et al. (2015). Apache Flink: Stream and batch processing in one engine. *IEEE Data Engineering Bulletin*.
[42] Han, J., & Stroulia, E. (2020). NoSQL data management for scalable analytics. *Journal of Big Data*.
[43] Noghabi, S., et al. (2020). The evolution of real-time big data architectures. *ACM Queue*.
[44] Gstrickland, J., et al. (2021). Hybrid analytics systems for real-time fraud detection. *Information Systems Frontiers*.
[45] Dutta, A., et al. (2022). Scalable architectures for machine learning in fraud detection. *Journal of Digital Banking*.