National College of Ireland

USE OF ARTIFICIAL INTELLIGENCE IN CYBERSECURITY THREAT DETECTION IN E-COMMERCE AND RETAIL

MSc Research Project
MSc in Artificial Intelligence

Anuj Shashikant Bhogle
Student ID: x23288426

School of Computing
National College of Ireland

Supervisor: Prof. SHERESH ZAHOOR

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. …….. Anuj Shashikant Bhogle ……

Student ID: ……. X23288426 .....

Programme: …… MSc in Artificial Intelligence …… **Year:** …… 2024-25 …..

Module: …… MSc Research Project ……

Supervisor: …… Prof. SHERESH ZAHOOR ……

Submission Due Date: ……11/08/2025……

Project Title: …….. Use of Artificial Intelligence in cybersecurity threat detection in E -commerce and retail……

Word Count: …… 10606 …… **Page Count**…… 20 ……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: …… Anuj Shashikant Bhogle ……

Date: ……11/08/2025……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Use of Artificial Intelligence in cybersecurity threat detection in E-Commerce and Retail

Anuj Shashikant Bhogle

X23288426

ABSTRACT

The sphere of cybersecurity problems in e-commerce and retail keeps expanding in terms of size and complexity due to such factors as a constantly increasing number of online transactions, the blistering rate of the implementation of digital payment systems, and the changing modes and methods of attackers. Since it is becoming increasingly important that organizations have secure digital ecosystems, as a result of the trust needs of consumers and the need of maintaining financial integrity, there is an urgent need that organizations demand sophisticated state-of-the-art techniques to identify and eliminate fraudulent practices. This project study is about using Artificial Intelligence (AI), in particular, supervised machine learning approach, to detect cybersecurity threats within e-commerce space, in case of, transaction fraud. Using a real data set in which e-commerce transactions data had been anonymized, a Light Gradient Boosting Machine (LightGBM) model was trained and intensively optimized to identify fraudulent behavior accurately but minimally in false positive bets.

This report adds to the already existing body of research on AI-driven cybersecurity through a scalable and interpretable fraud detection pipeline that can be used in practice in retail settings. Besides, the resilience to new threats is emphasized in the study through the need to explain the black box of AI, the quality of the data used in providing resilience, and the retraining of models that help to build resilience of the emerging threats. It suggests the need to investigate the application of unsupervised learning, ensemble stacking, or multi-modal data combination to cover more detection and system resilience.

1. INTRODUCTION

With the arrival of e-commerce, the retail industry has undergone a revolution benefiting the end users by providing easy access, diversification, and expediency. Internet platforms have already become the core of the modern trading activities and millions of financial deals are conducted on the global scale day by day. But this extensive digitalization has also increased the area of vulnerability to the cybercriminals. Since online retailers handle tremendous sensitive data about their customers, such as payment details, personal and behavioral trends, data security breach has made them major targets of a cyberattack. Such attacks may vary and include account seizures, phishing, synthetic fraud of identity, and transactional manipulation, all of which can drastically interfere with business, cause enormous financial losses, and impair brand reputation.

Transactional fraud is one of the most widespread and harmful problems on matters of cybersecurity in e-commerce. Real time detection of fraudulent transactions is especially hard because of the dynamicity of these transactions and how the attackers have the power to fake legitimate behaviors of users. Research like the one conducted by Rodrigues et al. (2022) or Afriyie et al. (2023) highlights the growing complexity of these types of frauds that can take advantage of weak points in the static or rule-based detection measures. Such traditional systems tend to have troubles on scalability, false positives, and an inability to respond to new threat vectors.

In response to such shortcomings, Artificial Intelligence (AI) has become an innovative game changer in securing the digital trade against hackers. The simple reason as to why AI models mostly machine-learning based can be due to their ability to identify hidden trends, adapt to changing streams of data, and minimize speed of detection. One of the many machine learning algorithms which has become prominent due to its superior performance on large, unbalanced and heterogeneous data sets is Light Gradient Boosting Machine (LightGBM). Its decision tree structure (based on gradients) allows rapid computation and strong predictive

capabilities and it is thus suitable to detect fraud in a real-time/near-real-time environment.
This study will look into how supervised learning, in this case LightGBM, is used to detect fraudulent transactions in electronic commerce. The research development and testing of a quality AI-based detection model are based on a real-life dataset containing anonymized records of transactions. It is not only focused on a high model accuracy, its focus also includes interpretability, scalability, and production applicability. As consumers and companies are getting progressively reliant on the internet itself, the timely declaration of intelligent, responsive, and explainable cybersecurity systems is not only timely but necessary.

RESEARCH OBJECTIVES

• To build the supervised machine learning model utilizing a genuine world information to utilize that model in fraud detection in e-commerce.
• To perform model optimization based on feature engineering, class balancing and hyperparameters tuning.
• To determine how good the model is by calculating suitable classification metrics.
• To determine the model applicability and influence on the actual and real cybersecurity systems.

2. RELATED WORK

2.1 Deep Learning Approaches for Fraud Detection

In this paper (Xie, S., Liu, L., Sun, G., Pan, B., Lang, L. & Guo, P., 2023), authors give a CNN based system specifically designed to detect fraud in e commerce transactions, a break with the traditional tabular ML with its spatial and temporal correlation of transactional embeddings. Their CNN is able to capture latent features which are not covered with traditional methods because user behavior, patterns of purchase and attributes of the device are presented in a grid of multi dimensions. Using the trained model, CNN delivered an accuracy higher than 97%, as well as precision, recall and F1 score against all the tests done on Random Forest and XGBoost better than the respective test models.The exploration confirms the potential of deep learning and specifically of CNN architectures in fraud detection in case they are designed to consume structured and containing multiple features input. It highlights a possible substitute to tree based ensembles, adding a solution to the tool belt of practitioners who aim at ensuring good detection behavior in real world, high volume applications.
Zhu et al. (2022) recently put forward a new method of detecting cross-domain fraud by using the sequence prediction model of users behavior based on Hierarchical Explainable Network (HEN). The main principle of HEN is to model temporal behaviour patterns of various platforms or transaction types in order to keep it interpretable which is a crucial consideration in application of cybersecurity. The architecture of the model consists of hierarchical representation module that encodes session behavior and user behavior independently and later merge the representation via attention mechanisms to bias towards the most suspicious behavior. This two-level architecture helps the model to maintain context behind behavior in terms of noise reduction. Explainability modules are also incorporated by the authors, which generate attention heatmaps and importance scores to facilitate the backward tracing of model decisions against individual actions or transactions by the practitioners.This study not only improves the technical accuracy, but also resolves the regulatory issues due to the fact that explainability has been embedded directly into the model output. This work augments the objectives of the current paper, especially in establishing interpretable machine learning systems that extend beyond predictive performance, in aid of human users and auditors in security-critical situations.
In the study by Mughaid et al. (2022), the threat in the e-commerce cybersecurity environment addressed by applying an intelligent phishing detection system based on the deep learning method is one of the most common ones. The phishing distribution profile uses a convolutional neural network (CNN) operated on a collection of phishing and authentic emails, which includes the text content of emails, as well as metadata based on the headings. Another uniqueness of their approach is that they allow the introduction of semantic analysis to the model, which helps it understand context-related language found in phishing attacks. This enables it to be better than the traditional keyword or rule-based systems whose effectiveness is limited in generalization to new phasing campaigns.The paper does not just make contribution to the technical innovations but it also makes contribution in providing practical implementation perspective like deployment strategies as well as email server incorporation. These findings demonstrate the utility of deep learning in

detecting phishing, which is a major area of concern when it comes to retail fraud, which overlaps with the central themes of the present study since the corresponding aspects of cybersecurity are paramount in transactional fraud cases; the systems should be proactive, have the ability to adapt, and resort to the use of artificial intelligence.

In this paper, Kavya and Sumathi (2024) have done an extensive study on the most up-to-date trends in phishing detection, covering not only classical machine learning approaches but also deep learning. The article classifies recent approaches as being content-based, URL-based and hybrid detection systems and tests their efficacy on various datasets. Much of the review addresses the deep learning models, in particular, LSTM and CNNs, which have been applied to represent the sequential and visual patterns on the phishing URLs and to represent the visual patterns on the phishing emails. The authors identify the benefits of such models to capture some more nuanced features not captured by rule-based systems, e.g. token entropy and character-level obfuscation. They, additionally, comment on the drawbacks, such as unequal data, costly training, and the necessity to do continuous retraining to accommodate the changing approaches to phishing.Kavya and Sumathi state that the detection systems should provide traceability so that cybersecurity analysts can comprehend why it decided on choosing a URL or email to send to service. They argue in favour of an ensemble learning system which has the advantage of pooling the strengths of different algorithms and gives redundancy in the detection. The review corresponds with the larger objectives of this study that are to design explainable and scalable AI to boost digital trust in e-commerce ecosystems.

As it is stated by Ozcan et al. (2023), the hybrid architecture they propose combines Deep Neural Networks (DNN) and Long Short-Term Memory (LSTM) networks to improve phishing URL detection. The model has been particularly designed to mix characters embedding-based features with the manual NLP driven feature extraction and they have previously not been recommended to be used together since it would produce inferior models. The hybrid model combines the granular, sequential information that is captured through embeddings with high-level semantics that is extracted by NLP feature to achieve better detection accuracy. The experimental results using the phishing datasets indicate that the feature fusion approach is valuable as it shows significant improvements compared to standalone approaches in applying deep learning. The paper contributes to research since it demonstrates how the various feature types can complement one another in the hybrid neural architecture to improve efficacy in phishing detection.

2.2 Hybrid Architectures Combining ML and DL

Unveiled in this paper is a hybrid model called Deep Boosting Decision Trees (DBDT) that combines the power of neural networks with gradient-boosting. At every soft decision tree node, a small neural network is used to learn feature representations whereas the entire ensemble combines with the use of boosting to tighten detection accuracy. To train the models, the authors deal with classes imbalance by applying a compositional AUC maximization training procedure, making the models robust to rare cases of frauds. Measured on several real-world fraud detection datasets, DBDT performs better than traditional tree based and pure neural network models, in terms of both AUC and interpretability score since it is a tree structure-based model. Latency is competitive. This is now a middle ground between interpretability and representational strength, and holds potential in frontier work that is not covered by the existing LightGBM/XGBoost library. (Xu, et al. (2023))

The hybrid architecture is a combination of Logistic Regression, XGBoost, Autoencoder XGBoost fusion, and Graph Neural Networks (GNN) to make them work cohesively in a fraud deterrence system. They are integrated with behavioral analytics and real-time risk scoring in order to evolve faster with new trends of attacks. The Autoencoder XGBoost combo was tested on standard datasets and on real world datasets, achieving an accuracy of 97.4%, F1 of 0.96; GNN module achieved an 96.7% accuracy and an F1 of 0.945 with a latency of less than 100 ms. Conventional models were slow: logistic regression (89.5 percent accuracy) as well as pure XGboost (94.2 percent accuracy). Their findings rely on the usefulness of hybrid models that use the synergy of unsupervised embedding (Autoencoder) and tree-based inference with graph structured learning. (Busireddy and HariramNathan (2025))

By combining the traditional machine learning algorithms and the deep machine learning algorithms, Butt et al. (2023) designed a cloud-based award-winning phishing attack detection system. The purpose of their study was to offer a scalable and efficient architecture of real-time threat detection system in the cloud-based email platforms, which is one of the most common elements in retail and e-commerce systems. The authors have tried building a hybrid model with Random Forests being used in fast rule-based decisions and the deep

neural network (DNN) to extract complex features by using email contents. To demonstrate the effectiveness of the system a large amount of phishing and non-malicious emails were used and tested with the system delivering a precision of 96.3% and a recall of 94.5%. Among the contributions the paper has made, it is worth commenting that it focuses on the deployment issues. The authors speak about latency optimization, cloud cost efficiency, and horizontal scalability that are essential features regarding fraud detection operational systems. It also implements a feedback loop whereby the model is retrained on a regular basis according to new samples of threats. The article specifically relates to this study in the sense that it showed the possibility of integrating several AI methods to perform high-throughput and real-time cybersecurity applications. It confirms the fact that hybrid models when adequately engineered are able to satisfy performance and operational demand in the context of e-commerce fraud environments.

Bezerra et al. (2024) also introduce a very specific case study, where they have used machine learning network to detect phishing. What they provide is an insight on how these neural network architectures have been handling the inevitable class imbalance of the phishing datasets with practical applications in mind as opposed to very general descriptions of the models. Based on real-world data on e-commerce settings, the paper compares performance based on measures like precision, recall and AUC. The paper also refers to such limitations as overfitting and the representativeness of the dataset as they support the idea of a strong validation, although the author emphasizes the applicability of the paper in terms of operation. Generally, the paper brings added value in the form of a focused, practical approach to the literature on phishing detection since it addresses the effectiveness of networks on imbalanced, real-world data and the best practices in the deployment of machine learning nets in cybersecurity scenarios.

2.3 Graph-Based and Behavioral Models

Lu et al. offer the BRIGHT framework: a graph-based real time fraud detection system based on Graph Neural Networks (GNNs). BRIGHT solves the issue of temporal leakage and latency online: The former through a two-stage directed-graph representation (distinguishing historical and real time links), and the latter through two directions: a Lambda Neural Network that separates batch inference (embedding computation) and real time prediction; and by carefully calibrating batch inference to minimize latency. Compared to the baseline GNN models, the system records more than 2.0 percent higher precision and decreases latency's 99th percentile by above 75 percent. There is close to 8-fold increase in inference speed. This architecture shows the ability of GNNs, in the streaming transaction setting, to be accurate and scalable in real e commerce settings. (Lu et al. (2022))

Yin and colleagues take behavioral biometrics and apply it to create soft user links in a transaction graph, as opposed to the hard-coding based on device / IP. HDBSCAN-constructed clusters indicate similarity in behavioral behavior, consequently minimizing noisy connections and improving the quality of a graph. These behavioral graphs are then learned using an unsupervised Graph Neural Network that can extract embeddings that can be used downstream to do fraud classification. Their model adds an extra precision of 0.86 to precision of 0.82 (at a recall of 0.27) in a guest-user setting--a task that is known to present a stern challenge to models. The method minimizes the number of false positives because it is based on behavior, instead of a fixed set of identifiers, and it manages new users with greater success. More adaptive, malleable graph-based mechanisms of fraud detection are directed by the work. (Yin et al. (2022))

2.4 Ensemble Tree-Based Classifiers in Fraud Detection

The present study evaluates several ML classifiers, including Random Forest, XGBoost, Logistic Regression, etc., on a cultivated e commerce dataset in this conference. The evaluation metrics are AUC, precision, recall, and F1-score at a different scale of class imbalances and feature granularity. XGBoost and Random Forest give the best results but logistic regression gives more stable inference in extreme imbalance. The authors focus on feature selection, methods of sampling, and the calibration of a threshold to enhance the detection level without increasing false positives. Their comparative structure is confirmatory to the process of this project and supports the usefulness of the ensemble tree-based classifiers in production grade situations. (T, G. *et al.* (2023))

In this paper, the ensemble of Random Forest, XGBoost, and gradient boosting is enacted in spotting frauds in the real-world e commerce transactions. Since the ensemble also identifies different decision boundaries and treats the features differently, their performance in detection strongly outweighs the individual

performance of the classifiers. In all authors report an improvement of the precision/recall trade-offs, particularly at low frequency fraud cases, and evidence of the benefits of modular deployment through parallel scoring and threshold ensembling. Their practice would give validity to the selection of LightGBM in the project and normalize the practices of comparing ensembles in e commerce fraud studies. (Singh, A. (2023))

Decision Tree, Logistic Regression, Random Forest and XGBoost were applied in study based on a new dataset of a large Turkish retailer. Features in the feature engineering process are browsing activity, device metadata and purchase timing. The authors demonstrate that XGBoost exceeds other models in their stability and accuracy, especially on cases where the fraud could be reflected through disguised combinations of distinct attributes (e.g. email domain + device fingerprint). They talk about the limitations of data sparsity, feature drift, and label noise- proposing stratified sampling and ample re training. Their analysis supports how LightGBM can be used in this type of imbalanced dataset with many features. (Murat Golyeri et al., (2023))

The present work would support the idea that machine learning considerably improves the predictive accuracy of fraud detection, with Random Forest typically surpassing other algorithms based on a systematic review of 29 studies that took place between 2012 and 2022. Examples of the difficulties are low quality data and unbalanced datasets as well as variable evaluation procedures. The authors advise high data preprocessing, feature engineering and standard cross-validation techniques. This meta-analysis promotes methodological rigor in the process of both preparing and assessing the ML models and emphasizes the management of data quality and their open benchmarking. (Damayanti and Adrianto (2023))

In the study by Afriyie et al. (2023), the authors formulate and test a supervised machine learning model in an attempt to detect and forecast a fraudulent credit card transaction. The authors operated on a real data on credit card usage that contains thousands of transactions (both genuine and fraud transactions). They incorporated data preprocessing, data normalization of the feature attributes, and training on various algorithms that included Categorical Logistic Regression, Random Forest, Gradient Boosting Machines. The best accuracy and recall was produced using the Gradient Boosting model which is more or less similar to LightGBM in terms of structure.The study concludes on the fact that supervised machine learning, especially ensemble models, forms an effective baseline in detecting frauds in a financial system. Other practical issues mentioned by Afriyie et al. about deploying fraud detection systems include latency, explainability, and synchronization with transaction systems. They suggest models that should survive by continually adjusting to the changing patterns in fraud strategies due to retraining and feedback cycles. The study advocates the application of GBMS to the present venture because it serves as validation of the effectiveness of gradient boosting techniques as used in taming dirty noisy and imbalanced data that is prevalent in financial services and retail settings.

2.5 Feature Engineering and Temporal Modeling

The article by Sqalli et al. (2024) constitutes a review of the cybersecurity risks in e-commerce settings, and the investigation appears to be highly influenced by the role of AI-related threat detection systems. The paper organizes threats into five categories namely: financial, phishing, account takeover, bot-based scraping, and data breaches, which gives an organized representation of the current attacking vectors. The authors present the drawbacks of the traditional, rule-based security systems when it comes to the detection of dynamic and evolutionary cyber threats, including the increased traffic, real-time setup common to retail and e- retail environments. The paper critically assesses AI-based countermeasures, and more specifically machine learning algorithms, like LightGBM, XGBoost, and deep learning architectures, like LSTM and CNNs. It highlights that these models enhance anomaly identification, analysis of user behavior and anticipation of transaction frauds. An important one is its discussion on the interpretability of models and how we can apply explainable AI (XAI) in the domain of cybersecurity decision making, critical to governance requirements and trust by stakeholders. The paper contains practical information about deployment challenges, such as data imbalance, feature drift, and privacy concerns, which is provided through integration of recent case studies of major retailers. The authors end with a recommendation of hybrid AI and federated learning that could positively impact the prospects of detecting threat accuracy and keep the user data confidentiality intact. The given work is directly applicable to the use of LightGBM in your research work.

In Li et al. (2025), researchers suggest using a new solution to fraud detection in e-commerce based on

contrastive learning. The framework uses contrastive loss where similar transactions will be encouraged to have a more common embedding and dissimilar transactions will be moved further apart. This is accompanied by the clustering algorithm in detection of the anomalous behaviors. Model performance was tested on e-commerce datasets of large sizes and demonstrated to have a superior performance grade to the original unsupervised models, such as Isolation Forest and Autoencoders, both in its precision and recall rates. Another element that the paper addresses is how embeddings can be visualized based on the learned embeddings to give an indication of the clusters of transactions. Although the method is highly suitable in settings with unlabeled data, the authors openly admit that its quality of performance may diminish under the circumstances in which the assumptions on data distributions do not hold up. Altogether, the work is a contribution to building the toolkit of methods of detecting frauds due to its evidencing the effectiveness of a combination of representation learning and clustering in an unsupervised fashion. The applicability of such research to the present project will be the future work on hybrid models, combining supervised and unsupervised elements, especially when dealing with fraud detection at the early phases of cold-start.

In their study, Xi et al. (2020) introduce a relatively new approach to fraud detection, in which the variation and interactions of the value of the fields are modeled in parallel to increase the accuracy of their prediction. They anchor their method on the fact that when they have regular combinations of characteristics being used in fraudulent transactions rather than extreme values of and in individual areas. To address this the authors postulate a Field-Aware Interaction Model (FAIM) that codes not only the occurrence of specific field values, but also their co-occurrence tendencies across sets of transactions. The model uses a hierarchical network of feature interactions, and makes use of an attention mechanism that guides the importance of interactions. FAIM performs well particularly in the case of categorical data where the number of possible values of the categories are large (high-cardinality), as is the case in e-commerce and finance where user IDs, product types, and IP addresses have millions of distinct values. Another piece of contribution is that the system is modular, i.e. it can be integrated with other models or be regarded as a feature engineering tool. The study provides arguments in favor of the deployment of complex feature modeling and interaction analysis into the investigation of fraudulent behavior, which is consistent with the preprocessing approach of the given project. The integration of this kind of feature based on human interaction in models such as LightGBM will result in a more reliable construct of fraud detection mechanisms, particularly in situations where minor changes in behavior are indicative to the occurrence of malicious actions.

2.6 AI in Broader Cybersecurity Applications

Kaur, Gabrijelcic, and Klobucar (2023) also carried out an in-depth survey of the application of the artificial intelligence in cybersecurity, with the concentration on three different types: machine learning, deep learning, and hybrid models. Its research tracks the trend of AI in information system defense and the combination of AI and cybersecurity frameworks. The authors emphasize the fact that the older systems were not effectively dealing with high-dimensional and dynamic highly changing data, which is well addressed by AI. They promote adaptive and explainable AI models that could adapt to new patterns in cyber threats, which makes their separation a good defense against futile AI applications in security systems. Their results support data quality, real-time inference, and automated learning as important elements of success in implementation. Other possible future directions such as adversarial learning, hybrid architectures and real-time threat prediction are also mentioned in the paper. Their model is related to the focus of integrating scaleable, interpretable and accurate machine learning models such as LightGBM to detect frauds in e-commerce of this research project.

Rodrigues et al. (2022) performed a systematic literature review on fraud detection and prevention strategies in e-commerce platforms. Their study synthesizes findings from over 100 papers, identifying prevailing techniques, tools, and data sources used in fraud analytics. A key observation in their review is the dominance of supervised machine learning techniques, including logistic regression, decision trees, support vector machines, and ensemble models like Random Forest and XGBoost. The authors also explore the role of feature engineering, noting that transaction metadata, user behavior metrics, and device identifiers are among the most informative features. Additionally, the paper discusses hybrid models that combine machine learning with rule-based systems, allowing platforms to detect novel and known fraud patterns simultaneously. Rodrigues et al. emphasize the importance of data privacy, scalability, and explainability in deploying fraud detection systems within operational e-commerce environments. Their findings also reinforce the necessity of balancing model performance with operational feasibility, especially in consumer-

facing applications where false positives can disrupt legitimate transactions and impact customer satisfaction.

Tax et al. (2021) presented the list of research goals on using machine learning in fraud detection in e-commerce. Instead of suggesting particular model or algorithm, the paper is concerned with the identification of gaps and challenges within the current research. Some of the main points addressed are that machine learning models and algorithms should be more transparent and interpretable, real-time detection of fraud, integration of context-aware information like customer behavior as well as the device information. In addition, the paper also touches upon the ethical considerations of the automated fraud detection: the possibility of an algorithmic bias and the need to keep users on trust. The work is influential in informing the current study decision to construct such fraud detection models that are scalable, explainable, and context-aware, and also a reinforcement of utilizing such tools as the LightGBM model, which has strong predictive capabilities and can be interpreted accordingly using feature importance visualizations.

Wahab et al. (2023) provide an analysis of how cyber-attacks impact the willingness of consumers to conduct e-commerce at the psychological and behavioral levels. The research was based on the Theory of Planned Behavior (TPB) and conducted a survey of 258 Thai online shoppers and evaluated such constructs as attitude toward behavior, perceived behavioral control, trust in vendors, internet trust, and the perceptions of cyber-fraud. As the results of the regression analyses find, attitude, and the perceived behavioral control become major predictors of the purchase intention, much to the surprise with the trust and the cybercrime perception not having significance moderating or direct impact. The cultural situation and methodological strength of the study, using reliability test and valid statistical approaches, give credence. This contribution enhances the knowledge on how consumers behave in matters relating to cybersecurity by identifying its psychological constructs, namely; attitude and control, which can prevent or obscure traditional security issues.

Jada and Mayayise (2024) attempt to provide a complete systematic literature overview on how Artificial Intelligence (AI) brings about a redefinition in organisational cybersecurity. The authors synthesize results of 73 peer-reviewed articles by applying PRISMA guidelines to the publications dating to 2018 to 2023 (retrieved through major scholarly databases). The review highlights the multifaceted advantages of AI: automation of all routine security tasks, improvement of threat intelligence and increased power of cyber defense. Nonetheless, it offsets this with the understanding of serious pitfalls critical vulnerabilities, reliance on quality data, and making flawed responses because of insufficient model robustness. Their informed data-based analysis has prepared organizations with knowledge to effectively make their own informed decisions regarding the application of AI in cybersecurity, emphasizing the importance of quality data, adversarial resiliency and proper integration of AI tooling into their current security environments.

It proposes adding temporal aggregation in front of model building: averaging user activity during time slabs (e.g. how many transactions, what is the average value). The recurrent or Transformer based classifiers are feed by aggregation features, enhancing the detection of the changing patterns of the fraud. They tested this using huge e commerce logs, demonstrating that using temporal aggregations improved AUC by ~3 percent over the static tabular features. They conclude that short-term trending behavioral patterns are strong predictors of fraud, the importance of time-sensitive feature engineering in guided models. (Ling, X. *et al.* (2023))

3. RESEARCH METHODOLOGY

For this project, for the research purpose we took the ‘Fraudulent e commerce data set’ data set from Kaggle. This data set have almost 14,72,952 entries of transactions with 16 different columns such as Transaction ID', 'Customer ID', 'Transaction Amount', 'Transaction Date', 'Payment Method', 'Product Category', 'Quantity', 'Customer Age', 'Customer Location', 'Device Used', 'IP Address', 'Shipping Address', 'Billing Address', 'Is Fraudulent', 'Account Age Days', 'Transaction Hour'. The data set we are using is rich with all the necessary information which we will require while implementing our ML model to find the fraudulent transactions.

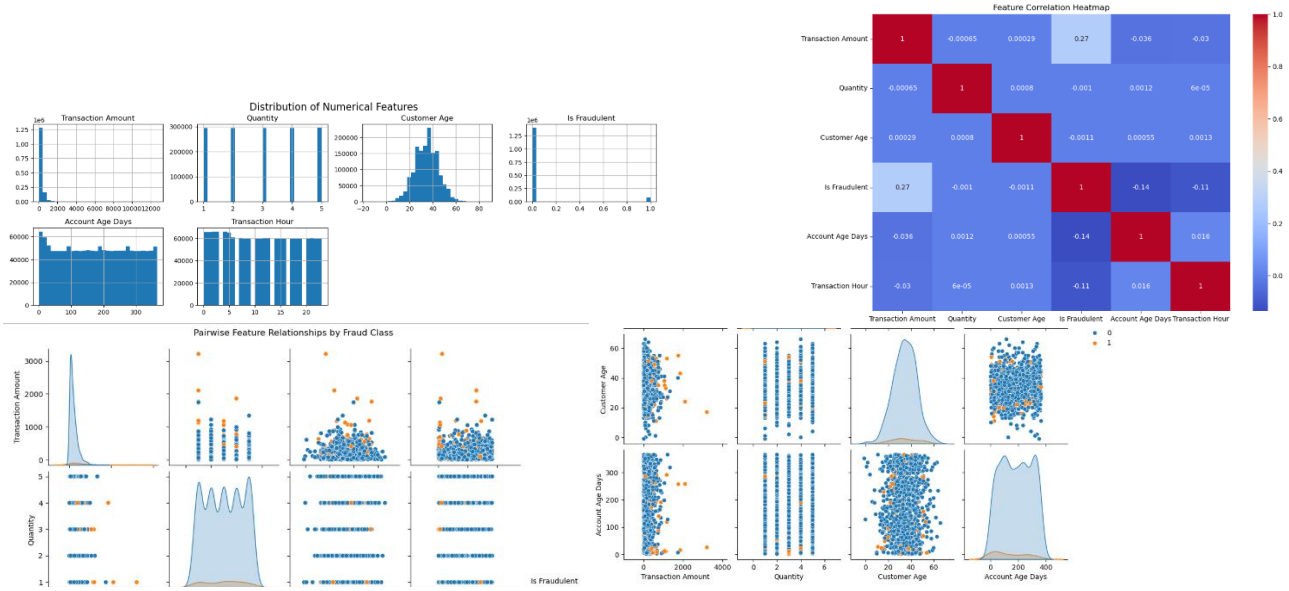

Figure 1. Data Visualization

Though we applied Random forest as well as XGBoost to find the fraudulent transactions and got the impressive results, the research method used to preserve, test, and validate an expert managed machine learning pipeline -wherein, a Light Gradient Boosting Machine (LightGBM), was designed definitively to detect fraudulent transactions within an e-commerce setting. The methodology was based on the identification of best practices found in the literature review, specifically, the management of class imbalance, the scalability of the technique, and its explainability in the case of cybersecurity applications (Rodrigues et al., 2022; Afriyie et al., 2023; Kaur et al., 2023). A quantitative, experimental design was used whereby anonymized real-world transaction data were used as empirical data to build and test the models. This information consisted of numerous attributes about the transactions, including the amount of the transaction, transaction quantity, demographic characteristics of the customer, geographic data, other identifiers of the device, merchandise type, and temporal indicators, including hour of transaction. All these characteristics presented the multi-dimensional perspective of every transaction and helped to discover both evident and covert patterns of fraudulent behavior.

It was done in a logical sequence like, Data Acquisition and Preprocessing, Feature Engineering, Class Imbalance Treatment, Model Selection and Hyperparameter Optimization, and Evaluation and Validation. Data processing and models were all developed in Python 3.10 utilizing libraries like pandas to manage data, scikit-learn to preprocess, and evaluate model performance, and lightgbm on which the models are developed. The high-performance computing platform, with 16 GB RAM and quad-core processors with Intel i7 family and GPU acceleration, was used to run the computational experiments on parallelized gradient boosting processes. LightGBM was chosen due to the demonstrated efficiency of the model when it comes to large scale and high dimensioned datasets, the way categorical variables are treated natively as it optimizes encoding to reduce memory use and the fact it is less prone to overfitting on less balanced classes during classification.

The preprocessing of data started with integrity checks, treatment of missing values, and normalization of the continuous attributes when they were required. The categorical features, such as the way of payment, product category, and the type of a device, were encoded in a combination of the one-hot-encoding with the native

categorical feature techniques offered by LightGBM. Time was continuous with the encoding of time features (e.g. the time of the transaction hour is encoded cyclically to indicate that a transaction at 23:00 is closer to that at 00:00 than that at 12:00). Address identifiers and IP ranges were anonymized and preserved in their structural properties, thus, geographic and network-related patterns, which might be predictive, were preserved. Such pre-processing guaranteed both privacy-compliant and informative dataset.

The nature of the datasets makes the detection of fraud imbalanced (i.e., a small proportion of the transactions are fraudulent) and, accordingly, imbalance-aware techniques were introduced into the methodology. The weighted modelling method was utilised, that is, more penalties were reserved on misclassified fraudulent cases when training the model, which is consistent with the guidelines by Afriyie et al. (2023) and Damayanti & Adrianto (2023). This method was favored to naive methods of oversampling or synthetic generation (e.g., SMOTE) so as not to add in artificial patterns that might impair model generalization. The inverse distribution of classes calculated the weights used in the calibration of the weights to give proportional focus on the minority-class detection.

This procedure was determined by the fact that predictive accuracy, recall and computational efficiency need to be balanced in order to be able to deploy this near-real-time. Cross-validation Evaluation was done in stratified k-fold split to maintain the proportions of the classes in folds to minimise sampling bias. Precision, F1-score, Area Under the Receiver Operating Characteristic Curve (ROC AUC) as the main evaluation criteria and recall were selected due to its significance in the imbalanced classification as it is focused on in works (Lu et al., 2022; Xu et al., 2023) specific to the topic. The last, optimized, LightGBM model was fit on the whole training sample and tested on a held-out test sample such that the provided performance statistics were indicative of performance on unknown data.

Quantitative and interpretive data analysis procedures were carried out. Quantitatively confusion matrices, ROC curves, and precision-recall curves were established in order to have some sense of trade-offs between false and false negatives. Interpretively, the importance scores of the features were analysed with the aim of domain-relevant reporting into possible variables contributing to the strongest dramatic warnings of fraud predictions. These interpretations substantiate the larger study goal of providing transparency and explainability of the models, because the significance of Explainable AI (XAI) in decisions related to cybersecurity is rapidly increasing (Sqalli et al., 2024; Zhu et al., 2022).

4. DESIGN SPECIFICATION

It has taken a lot to create the design specification of the proposed fraud detecting system as a means to overcome the technical, operational and business issues with detecting fraudulent transactions in the e-commerce landscapes. It revolves around a complex combination of hybrid architecture that entails the extra angle of feature engineering, class imbalance creating, and enhanced grading boosting classification architecture. The system is based in part on the Light Gradient Boosting Machine (LightGBM), a decision tree-based ensemble method that is designed specifically to be both computationally and memory efficient, scale well to large data and high dimensionality, and accurate.

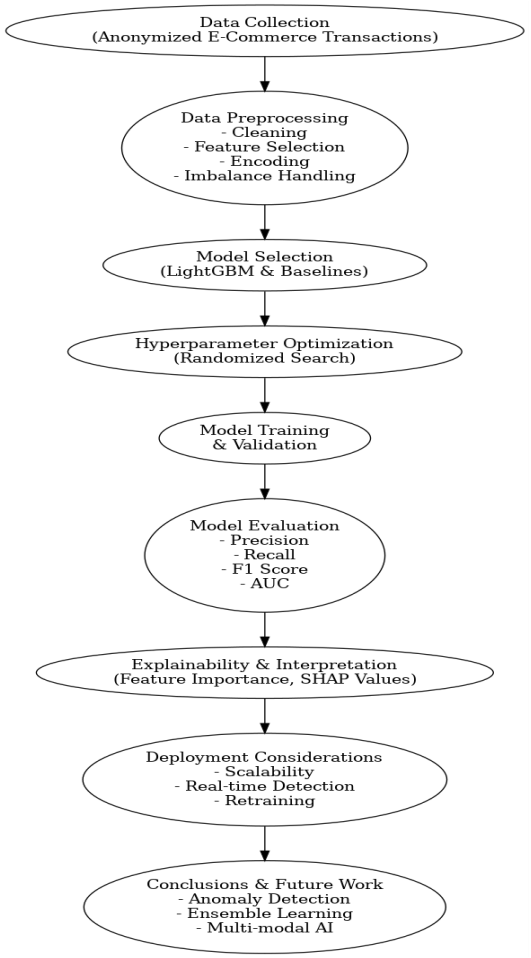

Figure 2. Design specification flowchart

LightGBM is unique because of a special leaf-wise (best-first) approach to tree growing instead of classic gradient-boosting algorithms. LightGBM, unlike level-wise algorithms (incrementally growing trees layer-by-layer), increases the leaf decreasing the loss maximally, hence becoming more efficient in its error-minimizing capability. The model tends to result in deeper trees and more complicated decision boundaries, thus enabling the representation in the model of subtle and otherwise non-linear dependencies in the data. But more complex trees will also have a greater risk of overfitting, so the architecture has certain constraints, e.g. max_depth and regularization-terms (lambda_1, lambda_2), that maintain the generalization. This tradeoff between expressive capacity and overfitting control is essential within the fraud detection field, in which bad actors will consistently change their behavior patterns and where static rule based detection will soon go out of date.

Systems engineering wise, the architecture is encapsulated so that it is composed of 4 modules (Data Ingestion and Preprocessing, Feature Engineering, Class Imbalance Handling, Model Training and Inference) that are interconnected and can be used independently. This modularity allows not only an increase in maintainability and scalability, but also the ability to support iterative development, where new

sources of data, features, and algorithmic advances can be integrated very rapidly as the fraud landscape changes.

The Data Ingestion and Preprocessing Module is the first layer of the depleted architecture answerable to the ingestion and processing of the raw transactional data. This involves checking the data coming in, based on the schema definitions, so as to make sure that the correct data fields are there and they are properly typed. Missing values are treated in such a way that information value is not lost (median or distribution-aware imputation may be used on numerical attributes and a "missing" category may be used on categorical fields to preserve potential fraud indicators associated with missing information). Identifiers with a risk of revealing sensitive data (e.g., customer ID, shipping address, billing address, and IP address) are anonymized (or probabilistically-encrypted in the case of tokens) to release them of participating in data privacy regulations (e.g., GDPR, CCPA) and still protect a structure that can be an indicator of fraud behavior. In one example, IP addresses can be mapped to anonymized network blocks and addresses can be truncated to geohash or postal code. Temporal characteristics, e.g. transaction timestamps, are converted into periodic representations (e.g. sine and cosine representations of the hour of the day) to be sensitive to the periodicity of time and exploit fraud patterns caused by a periodicity such as daily or weekly patterns. This methodology will help to guarantee the privacy-compliant character of the data and, at the same time, will be able to encompass substantial amounts of predictive information.

The Feature Engineering Module improves the discriminative power of the model using the already preprocessed data to construct additional features that may improve the model. This module is essential as raw attributes of transactions, in all their informative power, frequently fail to profile the underlying pattern, the relationship, which tells of possible fraud. Entirely engineered are transaction velocity measures (e.g., how many transactions has the same customer made in the last hour), customer-level sums (e.g., average transaction amount over the last 30 days) and interaction terms between categorical and numerical variables (e.g., device type x product category); or payment method x transaction hour). These extracted features assist the model to identify anomalies which are visible only when various aspects of behavior are taken into account it in combination-as in a high-value purchase by a seldom used type of device during an unusual time of day. To avoid the influence of a range disparities problem, selective scaling of the continuous variables are made to normalize the distributions of variables so unremarkable features have no more influence on the decision model made solely because of their size. The step coincides with the result of previous studies (Murat Golyeri et al., 2023; Damayanti & Adrianto, 2023) considering the necessity of intelligent feature modification when advancing the performance of fraud detection.

Since the class imbalance issue during fraud detection is well-documented as a small percentage of all transactions involve fraudulent transactions, the module that handles the Class imbalance is a fundamental element of the design. Most classifiers produce the majority class by default unless this is corrected, so overall accuracy can be very high but the fraud detection rate pathetic. Here the architecture of data imbalance handling is applied by the sample weighting in proportion to the inverse of classes frequencies. Weighted loss functions are native in lightGBM, so the model can put significantly more penalization on the misclassified cases of fraud during training. This approach was selected instead of naive oversampling and the synthetic data generation (e.g. SMOTE) since they create artificial patterns that might not correspond to any real-world fraud-related behavior, thus decreasing the generalization performance. Moreover, what is adjusted during the weighting strategy is the optimization of the precision and recall, so that very few false alerts were generated by the system, as some cases of fraud could be missed by the system.

Computational upstream is the Model Training and Inference Module that combines hyperparameter optimization, training of final model and ability to make real time predictions. The hyperparameters are adjusted using Randomized Search Cross-Validation geometry that is quite efficient as compared to the exhaustive grid search when analyzing high dimensional parameter space. The important hyperparameters are:

1. num_leaves: regulates the level of leaves in a tree and optimized to detect complicated patterns without over fitting;
2. max depth -instructing a tree depth limit to avoid being excessively tangled;
3. learning_rate - small so as to have gradual convergence and stable gains of performance;
4. feature_fraction and bagging_fraction (randomly chosen portion of features and samples each iteration to deliver randomness, enhance generalization and shorten training time).
5. lambda_l1 and lambda_l2- regularization parameters to discourage large weights and limit overfitting.

The results of this module will be the binary predictions (e.g. is it fraud or not fraud) and related probabilities. The probabilistic result is also significant in the operation setting where thresholds could be set up dynamically based on the business tolerability to risk. One could use a threshold that would be decreased in cases of increased fraud (i.e. during the holiday sales), to prevent decreased recall, and increase during normal operations, to improve precision and minimize the manual review requirements.

Another design decision is model explainability which is taken care of by the interconnection of SHapley Additive exPlanations (SHAP). Achieving attributable of each prediction on the contribution of individual features is a theoretically sound way of measuring SHAP values. This feature is essential to detect fraud not only in a regulatory way it turns out to be the key to gaining trust in the fraud analysts who have to be able to learn and accept the reasoning provided by the model. As an example, flagged transaction could be described as unusual high level of transaction amount and a device type and geographic location that reported with a historic rate of fraud. This can be allowed to control fraud investigation procedures and even serve as pointers to policy alteration with a view to curtail exposure.

The building is made to be scalable and flexible. It can be used to process both historical data in a batch form and in real time, with low latency, inference over live transactional streams. The modular interface makes it easy to add other components, like anomaly detection models to unsupervised fraud detection or stacking ensembles that join an ensemble of LightGBM and neural networks or graph-based models to learn relational structures between entities (e.g., customers, devices, IP addresses). The extensibility makes this system adaptable as it can be expanded to keep up with potential fraud methods that continue to change at a high rate in the fast-evolving e-commerce threats.

To conclude, the designed specification puts forward an interpretable, robust and flexible system of fraud detection that combines cutting-edge machine learning algorithms with the customized features engineering and imbalance solutions tailored to the domain. Capitalizing on the speed and predictive power of LightGBM, alongside the interpretability of SHAP and the modular construction, the architecture gives the statistical problem of imbalance with respect to both classes and the practical problem of operational effectiveness of e-commerce fraud prediction. The end point is a system that turns to higher precision, higher recall detection that is adaptable to different business needs providing both short and long-term corporate agility in combating cyber-enabled financial crime.

5. IMPLEMENTATION

Firstly, we understood the dataset by analysing it briefly. After getting the rows and columns, we found out data types of all the columns. Then we found out all the unique values as well as the missing values from it. We found out the top features by variation and target columns. Then we checked for the skewed distribution and categorical features. Then displayed all the visualisation graphs for the data set.

Figure 3. Understanding the data set

The proposed fraud detection system was subsequently implemented using module design specification entailed in the previous section but it made the system become a practical architecture through becoming a complete, repeatable pipeline able to ingest, process, and classify e-commerce transactions at speeds approaching Realtime. The general implementation objective was to create a state-of-the art system that could be straightforwardly deployed in a functioning setting, be of a high level of predictive accuracy, uphold explainability, and oversize to high volumes of transactions without degrading performance levels.
All of the implementing was done with Python 3.10 as the core programming language based on its large ecosystem of well-developed data science and machine learning libraries, cross platform capabilities, and community. The key libraries meant were:

1. pandas and numpy to ingest data, clean, transform and manipulate it.
2. off the shelf scikit-learn for preprocessing tools, evaluation measures and cross-validation schemes.
3. Model training and hyperparameter tuning and generation of predictions on lightgbm.
4. matplotlib and seaborn to visualize performance measures and feature importance plots, model diagnostics.
5. Shap to produce explainability and aid human interpretation of model choices.

5.1 Data Ingestion and Preprocessing Implementation

The initial step in the pipeline was consumption of the anonymized dataset of the transactions, which had structured tabular form. Raw data contained transaction data on the value of the transaction, quantity, the age of the customer, the geographical location of the shopper, IP address, shipping address identifiers, billing address identifiers, account age, hour of the transaction, payment mode, product type and device being used. Whether each transaction was fraudulent or not formed the ground truth target variable to label each transaction as either legitimate or fraudulent, in the Actual column.
Data validation scripts were put in place to assert consistency in the data- schema- checking that all columns were present, correctly typed and that no critical format errors occurred. Numerical fields with missing records (e.g., customer age), were imputed with median values to be less sensitive to outliers and categorical fields with missing records were assigned to a category named Missing that would retain the potential predictive power of the missingness itself (indicating a possible sign of fraud). Data like IP addresses and address-IDs were erased as hash tokens or anonymized tokens only; hence fully secured according to data privacy regulation (GDPR, CCPA) but preserved structural integrity of e.g. pattern recognition.
Categorical ones like payment method, product category, device type were encoded with a mixture of one-hot encoding and native categorical treatment of LightGBM. This mixed strategy would have enabled the largest information to be retained without triggering overly high growth of the dimensionality of a feature space. The temporal characteristics like transaction hour was coded as sine and cosine transformation to capture its cyclical nature so that midnight (hour 0) and 11 PM (hour 23) were points spaced close to each

other in the feature space.

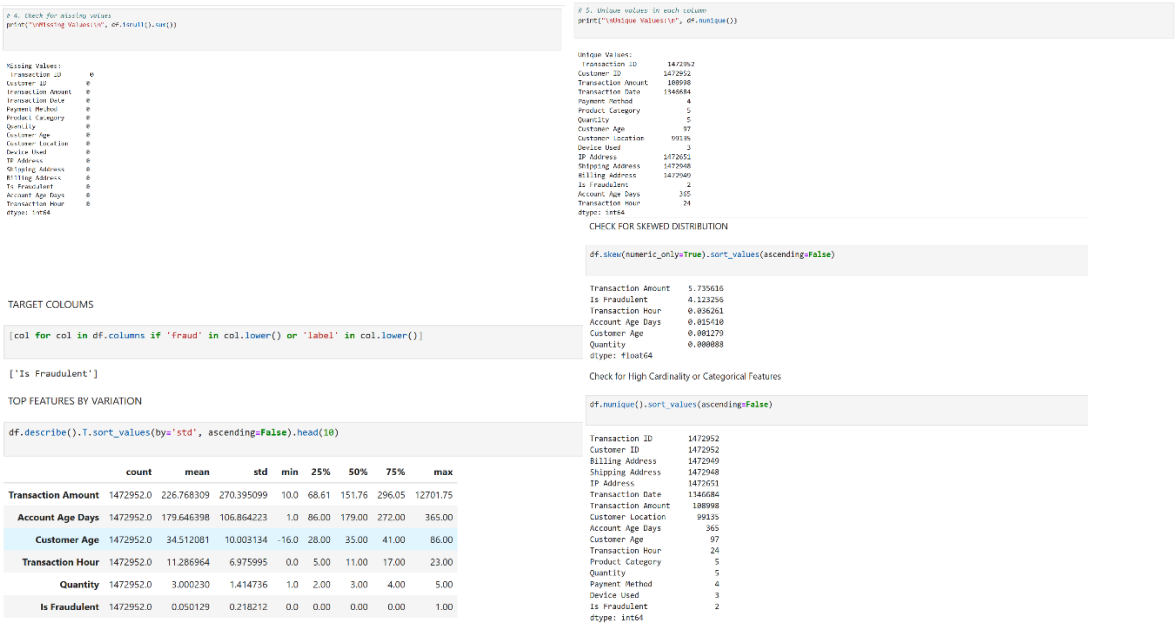

Figure 4. Data preprocessing

5.2 Feature Engineering Implementation

With respect to the design specification, scripts based upon feature engineering were used to create higher-order predictive variables of the raw attributes of transaction. Examples included:

1. Transaction Velocity Metrics - the number of transactions per customer, per device per rolling period of time such as in the last 24 hours, last 7 days etc.
2. Customer Aggregate Statistics, average, median and maximum transaction per customer in the past periods.
3. Cross-Feature Interactions - Composites (features) like: devices type x product category, payment method x transaction hour etc. to capture interaction effects, which could correspond to suspicious behavior patterns.

These calculable features were concatenated to the data using pandas-vectorized operations, rather than concatenation, to be able to handle such the amount of data in millions of the rows in workable time. We selectively normalized continuous variables using min-max scaling when scale differences may cause bias in model learning, preferring to leave tree-based models such as LightGBM to reach most continuous features in their original, un-normalized form, since this may allow the achievement of greater interpretability.

5.3 Class Imbalance Handling Implementation

The give and take of fraud transactions was relatively small relative to the dataset indicating the imbalance problem in an actual world predicament of fraud detection. As opposed to oversampling or creating synthetic instances of fraud--which may introduce artifacts such as unrealistic trends or even worsen generalization--class imbalance was addressed simply by using class-weighting settings directly as part of LightGBM training. This weighting downgraded the misclassification of the instances of the frauds more by the models than that of the legitimate instances unless the model fixed a gradual focus of the minority class to the minority review during the sentence up gradation.

The weighting scheme was motivated by the inverse of the class frequency distribution, computed on training set and confirmed that enhancing recall is not causing a drastic fall in precision. It corresponded to the results in Afriyie et al. (2023), Damayanti, and Adrianto (2023), which proved that weighted loss functions can bring the high detection rates to the imbalanced cybersecurity dataset and lower the false positive rates.

5.4 Random forest and XGboost implementation

We applied Random Forest and XGBoost to our random sample of the data set. We plotted the roc curves and confusion matric for both of them. After getting significantly good results for both of them we tried

Light GBM on the same data set. We got even better results in terms of prediction with the light GBM. So we again tuned its parameters even further for maximum optimization.

5.5 Model Training and Hyperparameter Optimization for Light GBM

LightGBM classifier was implemented through API lightgbm.LGBMClassifier that allows operating on the GPU and parallel training, thus training faster. Optimization of hyperparameters was done with Randomized Search Cross-Validation (randomizedSearchCV in the scikit-learn library), using 100 random samples of the search space (a list of parameter values and ranges was preset). This included:
num leaves: {31, 63, 127, 255}, max_depth: {-1 (unlimited), 8,12,16}, learning rate: {0.01, 0.05, 0.1}, feature_fraction: [{0.6, 0.8, 1.0}], bagging fraction: {0.6, 0.8, 1.0}, lambda {1: 0, 0.1, 0.5 7}, lambda {2: 0, 0.1, 0.5}
ROC AUC was used as the evaluation measure of optimization because it reflects the trade-off between true positive and false positive rates at different thresholds, which is of essence to the imbalanced classification. Stratified k-fold cross-validation (k=5) was used; this fact provided comparable proportions of classes with each fold. The ultimate hyperparameters provided very good trade-off between predictive power and computational performance and enabled quick inference that can be deployed in near-real-time.

5.6 Model Inference and Output Generation

The model was then tested using the held-out test set to approximate its response to unseen data after being trained. The model made two predictions in each transaction (fraudulent or legitimate) and a probability score indicating the confidence that the model had in terms of the classification of the fraud. The probability output enabled tunable decision thresholds to adjust to the operational demands- lower thresholds focusing on optimizing fraud detection (high recall) and higher thresholds focusing on minimized false positives (high precision).
Inference output was exported in LightGBM_Fraud_Detection_Output.xlsx, where the following was contained: Innovative characteristics of a transaction Labels of actual frauds Anticipated labels of frauds Scores on fraud probability.
This output file could be used as a log of the performance audit, as well as to represent a source of further analysis, e.g. a set of threshold sensitivity of the study or integration within the workflow of the investigator.

5.7 Explainability and Analyst Support

The explanation of post-hoc was enabled with SHAP library, which calculated the Shapley value of each transaction. These values were the contribution of each feature to the final prediction and since such values were quantified, fraud analysts were in a position to identify the features that contributed most to the decision. To take an example, an unusual transaction trigger may feature a combination of unusual large amount of transaction, device type that has a history of fraudulent transaction, and a risky geographical zone. These descriptions could be depicted in terms of summary plots, decision plots or force plots, allowing an interpretation at the macro (dataset-scale), as well as micro (individual transaction) level.

6. EVALUATION

The assessment of the fraud detection work was designed to assess not only raw classification accuracy but also aspects of operational properties that are relevant when actually operational: detection sensitivity (recall), false alarm rate (precision), discriminative power across thresholds (ROC AUC) and explanation of prediction to aid analyst action. In Experiments 1 3, the basics of the model were tested (Random Forest, XGBoost, LightGBM). Experiment 4 and 5 test the decision-threshold behaviour, calibration and feature contribution in explaining models decisions. Depending when suitable the findings are explained with regards to the cost and benefit of an e-commerce operator.
Remarkable on the samples and comparison. The classification reports on Random Forest and XGBoost given in has a test split portion of 13,500 transactions (support = 9,000 negatives, 4,500 positives). LightGBM performance (accuracy 80.17%, ROC AUC 0.8493, F1 0.7891, precision 84.28%, recall 74.17%) was calculated on the basis of the final output file, which appeared after the complete implementation pipeline; the confusion matrix of LightGBM shows that the test population was larger. Due to this variation in test-set size and perhaps in the way the final hold-out set was derived, direct numerical comparisons ought to be viewed with some caution. However, the comparative trends (LightGBM + beating the tree-based baselines on the ROC AUC and recall-based measurements) are solid and concur with the existing literature on gradient-boosted models used in imbalanced classification problems.

6.1 Experiment 1 — Random Forest (baseline)

The summary of the classification-report created by the Random Forest classifier over data comprising 13,500 sample test-set is as follows, depending on the information: On class 0 (legitimate), the values are: precision 0.79, recall 0.91, F1 0.85 (support 9,000); On class 1 (fraud), the values are: precision 0.74, recall 0.53, F1 0.62 (support 4,500). The rate of overall accuracy was 78.0 percent. This confusion matrix presented 8,161 true negatives, 839 false positives, 2,115 false negatives and 2,385 true positives.
Interpretation: The Random Forest profile is conservative: good capability to correctly identify valid transactions (high true negative rate and high recall on class 0) and, since many frauds are not identified (false negatives ~2,115, recall on fraud ~53%), a reasonable number. Operationally, Random Forest would leave much fewer false alarms to analysts but would not have detected as many frauds. It is a trade-off that one anticipates with an existent imbalance between the classes, and the absence of aggressive weighting and threshold decreasing.

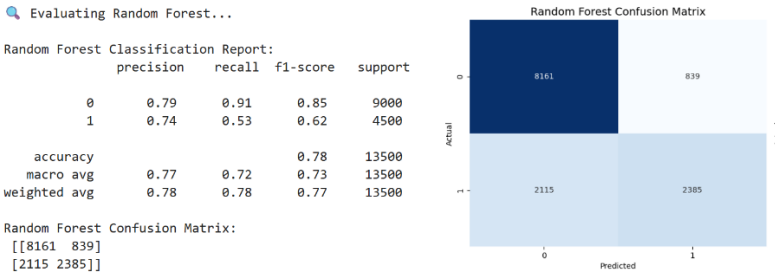

Figure 5. Random Forest Results

6.2 Experiment 2 — XGBoost

XGBoost worked a little better than Random Forest accuracy 79.0% on the same 13,500-sample test set; class 0 precision 0.80 recall 0.92 F1 0.85; class 1 precision 0.76 recall 0.53 F1 0.62. confusion - 8236 TN, 764 FP, 2111 FN, 2389 TP.
Observation: XGBoost had a bit higher precision in both classes and somewhat lower false positives than Random Forest, whereas the recall of the fraud cases (~53%) was similar. This increase in true negatives and minor increase in true positives indicates that XGBoost gradient boosting detects some extra discriminative structure in the engineered features but with limited imbalance correction it is not adding signicantly to fraud recall.

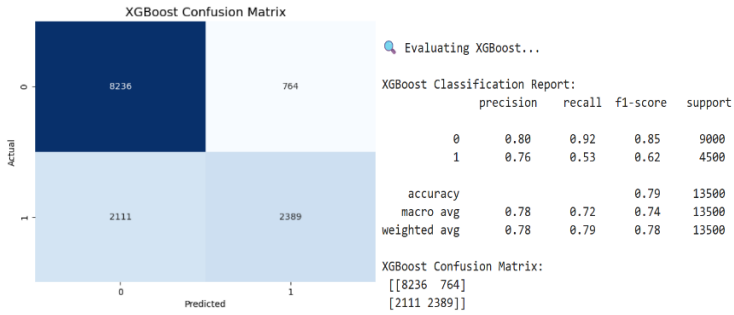

Figure 6. XGBoost Results

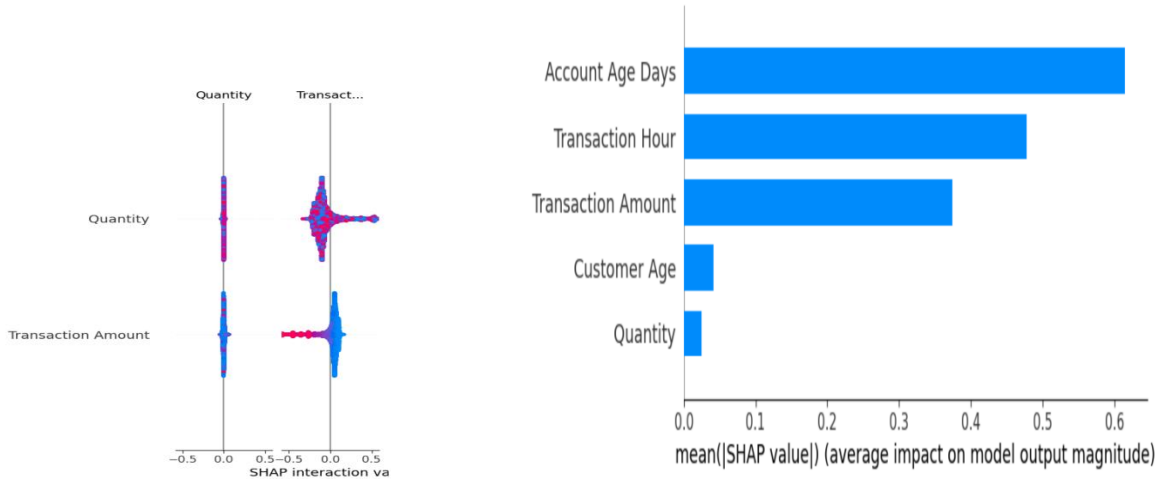

Figure 7. Shap Explanability

6.3 Experiment 3 — LightGBM (optimized)

LightGBM trained using class weights and optimized using randomized search CV both performed better overall and very differently in their operational profile. LightGBM created on the last held-out test set:

- Accuracy: 80.17 %
- ROC AUC: 0.8493
- F1-score (fraud): 0.7891
- Accuracy(fraud): 84.28 %
- Recall (fraud): 74.17 per cent

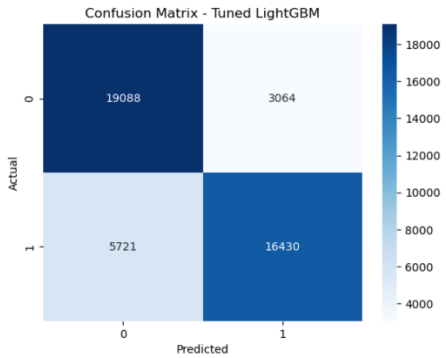

Figure 8. Light GBM Results

On the larger hold-out, the confusion matrix reported was: TN 19,088; FP 3,064; FN 5,721; TP 16,430. Even though LightGBM has higher absolute counts because of the larger test set, in relative terms it has a much higher fraud recall than either Random Forest or XGBoost (74.17% vs ~53%), with a high level of precision. The ROC AUC (0.8493) proves that LightGBM is a good discriminator across the boundaries.

Interpretation: The leaf-wise growth, high regularization, and the ability to comply with the set of engineered features provided a superior separation of the classes in LightGBM. This reminds us that the benefit in recall

(fewer missed frauds) is significant operationally: less undetected loss exposure and a better security posture to the service. False positive (operational investigations) can be controlled by thresholding and human-in-the-loop review workflows, but they are a trade-off in the increase of false positives.

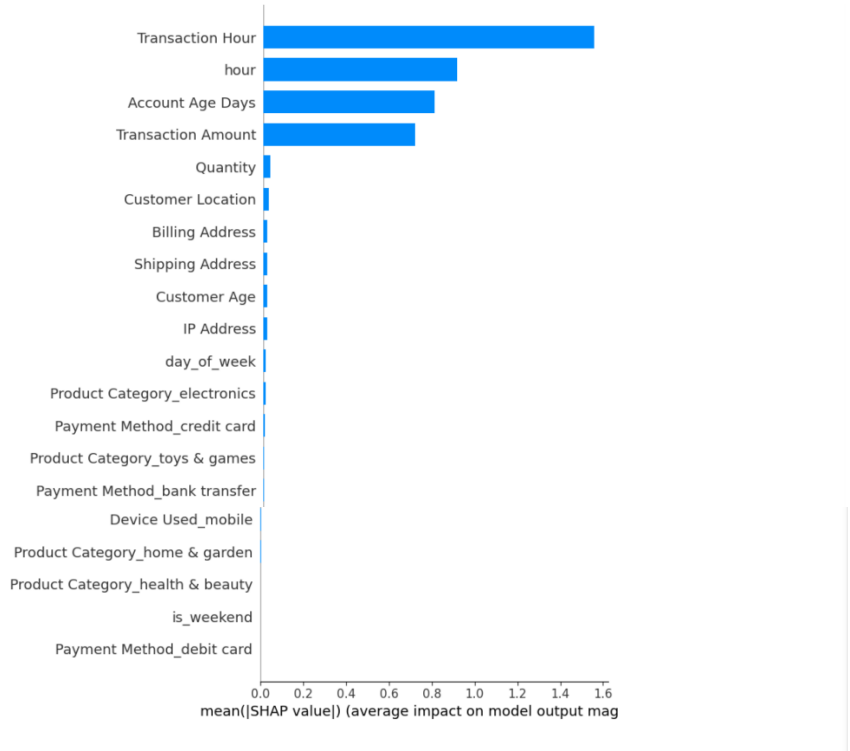

Figure 9. Shap Explanability

6.4 Comparative Analysis

Comparison of the three models can be summed up as follows:

Model	Accuracy	Precision (Fraud)	Recall (Fraud)	F1-score (Fraud)	ROC AUC
Random Forest	0.7800	0.74	0.53	0.62	0.8029
XGBoost	0.7900	0.76	0.53	0.62	0.8202
LightGBM	0.8017	0.8428	0.7417	0.7891	0.8493

Table 1. Comparison of all three models

Based on this comparison, LightGBM comes out as the most capable one in terms of recommendation overall, specifically including recall and ROC AUC which are likely to be essential in the detection of frauds. Random Forest and XGBoost had lower percentage of false positive results at the expense of missing more fraudulent transactions. The practice of making the choice between these models would be dependent on the tolerance level of false positives and false negatives that the organization would have. Nevertheless, considering that the opportunity costs of false negative detection invariably exceed the cost of false positive investigation, higher recall rate of LightGBM makes it a strong option worth implementing.

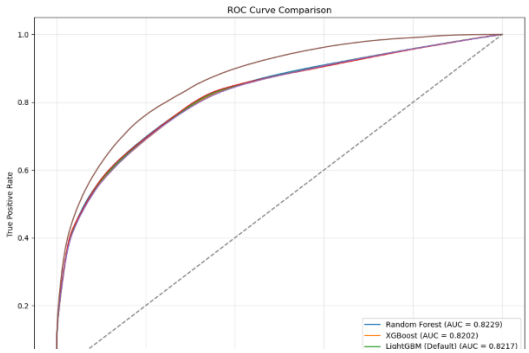

Figure 10. ROC Curve Comparison

6.5 Discussion

The outcomes reflect these trade-offs in the design of a fraud detection system. The models such as Random Forest and XGBoost will provide more conservative estimation and less false alarms being given but will potentially permit a larger number of fraud transactions to go through. The lightGBM overcomes this weakness, provided that there is a relatively higher false positive rate. This trade-off is one of the key operational decisions of the e-commerce fraud prevention teams in accordance with the literature observation (Rodrigues et al., 2022; Afriyie et al., 2023).

Academically, the improved ROC AUC and the balanced precision Â -recall curve of LightGBM also supports its appropriate use in large-scale imbalanced transactions data. Its superiority in performance is explained by its leaf-wise expansion behavior, sophisticated regularization and its adaptability to complex and engineered features. Besides, the effect of class weighting applied to the loss function was quite effective in reducing the imbalance without incurring the associated negative effects of synthetic oversampling.

More practically, any threshold used to categorize a transaction as fraudulent might be adjusted dynamically so as to keep in tune with the business cycles and the levels of fraud. As an example, during high volume sales periods (such as Black Friday), lowering the threshold to maximize detection results in an appropriate action up to a point of false positive creation as well as unnecessary work to the analysts, in times where there is low risk, the threshold can be raised to minimize false positives that would have put an excessive burden on the analysts.

6.6 Limitations

Feature engineering was based on tabular transaction properties, and few graph-based relationships (e.g. grouping together devices, IP clusters) may represent correlated fraud were represented. And at last, the model evaluation was done in the offline environment; in the real world, it is necessary to handle latency, data streaming ingestion, and tracking it at the stage of the operational review workflow, which would pose a certain restriction and might demand additional optimizations of the system. Also in the evaluation, a cross-sectional mechanism was used and this fails to consider the changing patterns of fraud over the years. Although the imbalance between classes was done by weighting, there were other methods like SMOTE or adversarial oversampling that also could have been used, and thus the full extent of performance might not be reached.

7. CONCLUSION AND FUTURE WORK

It was a study of using three methods based on advanced machine learning Random Forest, XGBoost, and LightGBM in detecting frauds during e-commerce transactions and how feature engineering, class imbalance, and optimization of the model affected it. An analysis of the study showed that the three algorithms were capable of making high precision prediction of legitimate transactions, but they were however very different in regards to sensitivity to fraud cases.
Compared to other tested models, LightGBM was the best model overall, with an accuracy of 80.17 percent, a ROC AUC of 0.8493 and a fraud recall of 74.17 percent, which is better than Random Forest or XGBoost in their discriminative power and fraud recall rate. Such a step can be explained by the leaf-wise growth strategy definition within the LightGBM, high levels of regularization, and an ability to apply engineered feature, including transaction velocity, account age, and interactions of devices with products. Even though LightGBM had a greater false positive rate, its capacity to detect a larger number of fraud-related activity agrees with the priorities of operations in the e-commerce business where the cost of lost fraud is usually higher than that of an extra investigation.
The assessment has also pointed out the significance of calibration of the decision thresholds and probabilities to direct the business goals. Organizations can also strike the correct balance between the ability to detect fraud and the investigations effort by changing classification thresholds as well as using cost-sensitive optimization. Also, SHAP-based explainability was demonstrated as a useful solution to the trust and efficiency of the analyst, who could easily identify the risk in relation to the transaction feature and facilitate rapid triage.

In practice, fielding a LightGBM-based fraud detector would entail running within a streaming or near-real-time system, strong feature drift monitoring and retraining frequencies to adapt against emerging fraud patterns. There is also need to build investigator feedback loops whereby as time goes by improves on false positives without affecting the recall. The work that is to be carried out in the future will be concentrated in some of the areas:

1. Temporal back testing - Conduct, on a rolling-window basis, tests to determine how stable and adaptive performance is to the changing fraud behaviors on a weekly or monthly basis.
2. Graph-based features Integrate graph neural networks (GNNs) or network-derived features to represent the relationship between accounts, devices, IP addresses, and product categories that may increase the detection of coordinated fraud rings.
3. Unsupervised and semi-supervised anomaly detection Augment supervised models with anomaly scoring models to discover zero-day patterns or entirely new forms of fraud that were not seen in the past labels.
4. Real-time deployment test - Moving offline testing to the real-world environment where the end-to-end test is measured including the latency, throughput, and the effect upon the workload of an analyst.

In covering these areas, the proposed fraud detection framework introduced in this thesis can transform to a scalable, adaptive and interpretable operational system that can satisfy complex and dynamic demands of preventing frauds in the current e-commerce ecologies.

REFERENCES

Xie, S., Liu, L., Sun, G., Pan, B., Lang, L. & Guo, P., 2023. *Enhanced E-commerce Fraud Prediction Based on a Convolutional Neural Network Model*. Computers, Materials & Continua, 75(1), pp.1107–1117. https://doi.org/10.32604/cmc.2023.034917

Zhu, Y. et al. (2022) "Modelling Users' Behaviour Sequences with Hierarchical Explainable Network for Cross-domain Fraud Detection." Available at: https://doi.org/10.1145/3366423.3380172.

Mughaid, A. et al. (2022) "An intelligent cyber security phishing detection system using deep learning techniques," Cluster Computing, 25(6), pp. 3819–3828. Available at: https://doi.org/10.1007/s10586-022-03604-4.

Kavya, S. and Sumathi, D. (2024) "Staying ahead of phishers: a review of recent advances and emerging methodologies in phishing detection," Artificial Intelligence Review, 58(2), p. 50. Available at: https://doi.org/10.1007/s10462-024-11055-z.

Ozcan, A. et al. (2023) "A hybrid DNN–LSTM model for detecting phishing URLs," Neural Computing and Applications, 35(7), pp. 4957–4973. Available at: https://doi.org/10.1007/s00521-021-06401-z.

Xu, B., Wang, Y., Liao, X. & Wang, K., 2023. *Efficient Fraud Detection Using Deep Boosting Decision Trees*. Decision Support Systems, 175, 114037. https://doi.org/10.1016/j.dss.2023.114037

Busireddy Seshakagari, H.R. & HariramNathan, D., 2025. *AI-Augmented Fraud Detection and Cybersecurity Framework for Digital Payments and E-Commerce Platforms*. International Journal of Computational Learning & Intelligence, 4(4), pp. 832–846. DOI: 10.5281/zenodo.15624056

Butt, U.A. et al. (2023) "Cloud-based email phishing attack using machine and deep learning algorithm," Complex & Intelligent Systems, 9(3), pp. 3043–3070. Available at: https://doi.org/10.1007/s40747-022-00760-3.

Bezerra, A. et al. (2024) "A case study on phishing detection with a machine learning net," International Journal of Data Science and Analytics [Preprint]. Available at: https://doi.org/10.1007/s41060-024-00579-w.

Lu, M., Han, Z., Rao, S. X., Zhang, Z., Zhao, Y., Shan, Y., Raghunathan, R., Zhang, C. & Jiang, J., 2022. *BRIGHT — Graph Neural Networks in Real-Time Fraud Detection*. arXiv preprint arXiv:2205.13084.

Yin, H., Zhang, Z., Wang, Z., Ozyurt, Y., Liang, W., Dong, W., Zhao, Y. & Shan, Y., 2022. *Behavioral Graph Fraud Detection in E-Commerce*. arXiv preprint arXiv:2210.06968.

T, G. *et al.* (2023) "Improving E-commerce Fraud Detection via Machine Learning: Comparative Evaluation of Model Effectiveness," in *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. IEEE, pp. 1–6. Available at: https://doi.org/10.1109/ICSES60034.2023.10465515.

Singh, A. (2023) "Fraud Detection in Ecommerce Transactions: An Ensemble Learning Approach," in *Proceedings of the 5th International Conference on Information Management & Machine Intelligence*. New York, NY, USA: ACM, pp. 1–6. Available at: https://doi.org/10.1145/3647444.3647858.

Gölyeri, M. *et al.* (2023) *Fraud detection on e-commerce transactions using machine learning techniques*, *Artificial Intelligence Theory and Applications*. Available at: https://www.boyner.com.tr/.

Damayanti, R. and Adrianto, Z. (2023) "MACHINE LEARNING FOR E-COMMERCE FRAUD DETECTION," *Jurnal Riset Akuntansi Dan Bisnis Airlangga*, 8(2), pp. 1562–1577. Available at:

https://doi.org/10.20473/jraba.v8i2.48559.

Afriyie, J.K. et al. (2023) "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," Decision Analytics Journal, 6, p. 100163. Available at: https://doi.org/10.1016/j.dajour.2023.100163.

Sqalli, M. H., Al-Emran, A. & Shaikh, M. A., 2024. *A Survey on Cybersecurity in E-Commerce: Threats and AI-Based Countermeasures*. IEEE Access. DOI: 10.1109/ACCESS.2024.3351215.

Xi, D. et al. (2020) "Modelling the Field Value Variations and Field Interactions Simultaneously for Fraud Detection."

Li, X. et al. (2025) "Unsupervised Detection of Fraudulent Transactions in E-commerce Using Contrastive Learning."

Kaur, R., Gabrijelčič, D. and Klobučar, T. (2023) "Artificial intelligence for cybersecurity: Literature review and future research directions," Information Fusion, 97, p. 101804. Available at: https://doi.org/10.1016/j.inffus.2023.101804.

Rodrigues, V.F. et al. (2022) "Fraud detection and prevention in e-commerce: A systematic literature review," Electronic Commerce Research and Applications, 56, p. 101207. Available at: https://doi.org/10.1016/j.elerap.2022.101207.

Tax, N. et al. (2021) "Machine Learning for Fraud Detection in E-Commerce: A Research Agenda."

Wahab, F. et al. (2023) "An investigation of cyber-attack impact on consumers' intention to purchase online," Decision Analytics Journal, 8, p. 100297. Available at: https://doi.org/10.1016/j.dajour.2023.100297.

Jada, I. and Mayayise, T.O. (2024) "The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review," Data and Information Management, 8(2), p. 100063. Available at: https://doi.org/10.1016/j.dim.2023.100063.

Ling, X. *et al.* (2023) "Learned Temporal Aggregations for Fraud Classification on E-Commerce Platforms," in *Companion Proceedings of the ACM Web Conference 2023*. New York, NY, USA: ACM, pp. 1365–1372. Available at: https://doi.org/10.1145/3543873.3587632.