ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb422

Design and Implementation of an Ontology-Based Fraud Detection System for Identity Theft in E-Commerce Transactions

Rebecca Ugonna Ugoha[1]; Chukwu Peter Chijioke[2]; Rosemary Chinyere Okolie[3]; Cosmas Ifeanyi Nwakanma[4]; Stanley Adiele Okolie[5]; Onyemauche[6]; Atumonyego Patricia Chiamaka[7]

[1;2;3;4;5;6;7]Federal University of Technology Owerri (FUTO)

Publication Date: 2026/02/28

Abstract: Identity theft in e commerce transactions results to a significant cybersecurity challenge with worldwide fraud losses exceeding 32 billion in 2022. Traditional rule based detection systems lacks flexibility while machine learning methods face transparency issues critical to regulatory compliance. This paper introduces an ontology based fraud detection framework using OWL 2 and SWRL to capture semantic relationships in e commerce transactions. The system encodes domain knowledge through 2 core classes and 23 properties creating understandable detection rules through semantic reasoning. Testing with synthetic data shows 91 percent precision and 88 percent recall with an average reasoning time of 5.4 seconds for 150 transactions. The framework identified 71 semantic axioms offering traceable decision paths for each classification. Although current scalability limits real time use this approach provides a foundation for explainable AI in financial fraud detection.

How to Cite: Rebecca Ugonna Ugoha; Chukwu Peter Chijioke; Rosemary Chinyere Okolie; Cosmas Ifeanyi Nwakanma; Stanley Adiele Okolie; Onyemauche; Atumonyego Patricia Chiamaka (2026) Design and Implementation of an Ontology-Based Fraud Detection System for Identity Theft in E-Commerce Transactions. *International Journal of Innovative Science and Research Technology*, 11(2), 2013-2017. https://doi.org/10.38124/ijisrt/26feb422

I. INTRODUCTION

The speedy growth of e Commerce reaching 5.2 trillion globally in 2021 and projected expansion to 8.1 trillion by 2026 has created unprecedented opportunities for cybercriminals. Identity theft involving unauthorized use of personal information for fraudulent transactions accounts for 43 percent of all reported fraud cases according to the 2023 Consumer Sentinel News Report.

Existing fraud detection systems predominantly employ two techniques the rule based systems with predefined thresholds and machine learning models using statistical pattern recognition. Rule based systems while understandable suffer from static rule set that fraudsters can learn to bypass. Machine learning strategies particularly deep neural networks achieve high accuracy but operate as black boxes limiting adoption in regulated financial environments where decision explanations are not optional.

II. RELATED WORK

The rapid growth of e commerce has greatly increased the number of online transactions which brought about the exposure to cybercrime particularly identity theft and transaction fraud (Statista, 2022). Identity theft which involves the unauthorized use of personal information has now become one of the most commonly occurring forms of fraud in e commerce causing financial losses, reputation damage and reduced consumer trust (Adamu et al., 2022).

Early fraud detection systems were almost entirely rule based it depends on expert systems and static limits marking transactions exceeding stated thresholds though interpretable but such methods generate high false positives due to their inability to adapt to the dynamic way of evolving fraud strategies (Brown Williams, 2020). To solve these problems machine learning techniques have been widely adopted to detect complex transaction patterns using Statistical models such as logistic regression and decision trees improved flexibility by considering multiple features but it often struggled with non-linearity (relationship that cannot be associated to a straight line) and interactions among many variables (Hasan, 2024 and Chowdhury, 2024).

Machine learning approaches have greatly advanced this landscape. Ensemble methods such as Random Forests, XGBoost, and boosted decision trees frequently lead benchmarks in credit card fraud detection and other financial fraud tasks (Damian Chukwujekwu et al. 2024). Deep

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb422

learning techniques especially recurrent neural networks (RNNs) and long short term memory (LSTM) models in combination with attention mechanisms are effective in modeling sequential transaction patterns and temporal anomalies. Recent hybrid ML + DL stacking models achieve F1 scores above 94 percent (Siddique, 2025). Graph neural networks (GNNs) heterogeneous graph models and transformer based modules have begun to capture relational patterns between users, devices, merchants and transactions. A 2025 study on heterogeneous GNNs with graph attention and temporal decay improves detection on real transactional networks by weighting different relation types and adjusting for timing effects (Ion Grujdin Datcu 2025, Ziyi Zhang et al. 2025). While machine learning models demonstrate improved detection accuracy they often require large labeled datasets and suffer from limited interpretability making them less suitable for transparent decision making in financial systems (Green Davis, 2020).

Ontology based and hybrid semantic systems offer complementary advantages they encoded domain knowledge, relationships and constraints are clearly defined facilitating reasoning and interpretability (Orche Bahaj, 2020, Kainat Ansar et al. 2025). Recent works include enhancing fraud detection in SWIFT financial systems through ontology based knowledge integration and graph driven analysis, which combine ontology with graph representations to model customer behavior. Another study involves the combination of ontologies and neural networks to detect payment system fraud, integrating semantic features to improve anomaly detection in structured transaction data (Ada John et al., 2025).

In spite of this ontology based approaches are still underused particularly for identity theft in ecommerce contexts. Key gaps include:

Few systems integrate ontological semantic reasoning with cutting-edge deep learning or graph models in a unified architecture (Li-Ming Chen et al., 2025).

Little work has addressed identity fraud specifically in e-commerce capturing identity verification data, device fingerprints, behavioral anomalies and biometric or document fraud (Saritha Crasta Janefer, 2025).

Scalability of ontology maintenance and keeping semantic knowledge up to date with novel fraud strategies are often not addressed (Ritesh Chandra et al., 2025).

Evaluation often uses static datasets real time detection and deployment considerations are less frequently studied (Otieno, 2025).

➢ *Research Gap & Motivation*

In summary existing ontology based solutions often focus on financial fraud and provide limited evaluation within e commerce identity theft. few studies shows end to end pipelines that integrate real transaction datasets semantic reasoning and automated fraud labelling although ML/DL/GNN techniques achieve great accuracy they suffer from a lack of interpretability and vulnerability. Ontology based systems bring transparency and semantic reasoning but have not yet been thoroughly combined with modern ML/GNN methods in the specific domain of identity theft in e commerce this gap motivates the development of an ontology driven fraud detection system specifically designed to identify identity theft in e commerce transactions with improved interpretability and adaptability.

III. SYSTEM ARCHITECTURE

Our ontology based fraud detection system comprises four integrated components: • Knowledge Modeling Layer (OWL 2 ontology).

- Rule Engine (SWRL).
- Semantic Reasoner (Pellet)
- Data Integration Module

IV. SEMANTIC RULE IMPLEMENTATION

Rules identify fraud patterns such as address mismatches, new account anomalies, suspicious device usage, and geographic inconsistencies. Each rule was implemented in SWRL and validated against the ontology.

V. EXPERIMENTAL SETUP

We constructed a synthetic dataset of 150 transactions with 23 attributes per transaction and a 26.7 percent fraud rate. The implementation used Protege 5.6.0, Pellet 2.3.6, and Python 3.9 with Owlready2. Evaluation was performed with five-fold cross-validation.

VI. PROTOTYPE APPLICATION AND TESTING

To validate the ontology in a practical environment, we implemented a prototype using FastAPI (backend) and Next.js (frontend). Users can upload Excel transaction datasets to the backend, where records are matched against ontology classes. Fraud and non-fraudulent cases are classified and returned to the front end in tabular format. WebSocket communication streams results to the frontend in real time, and operation trails are logged in the UI for transparency. A. System Architecture The prototype architecture is illustrated in Fig. 1, showing the data flow between the frontend, backend, ontology reasoning engine, and user interface.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb422

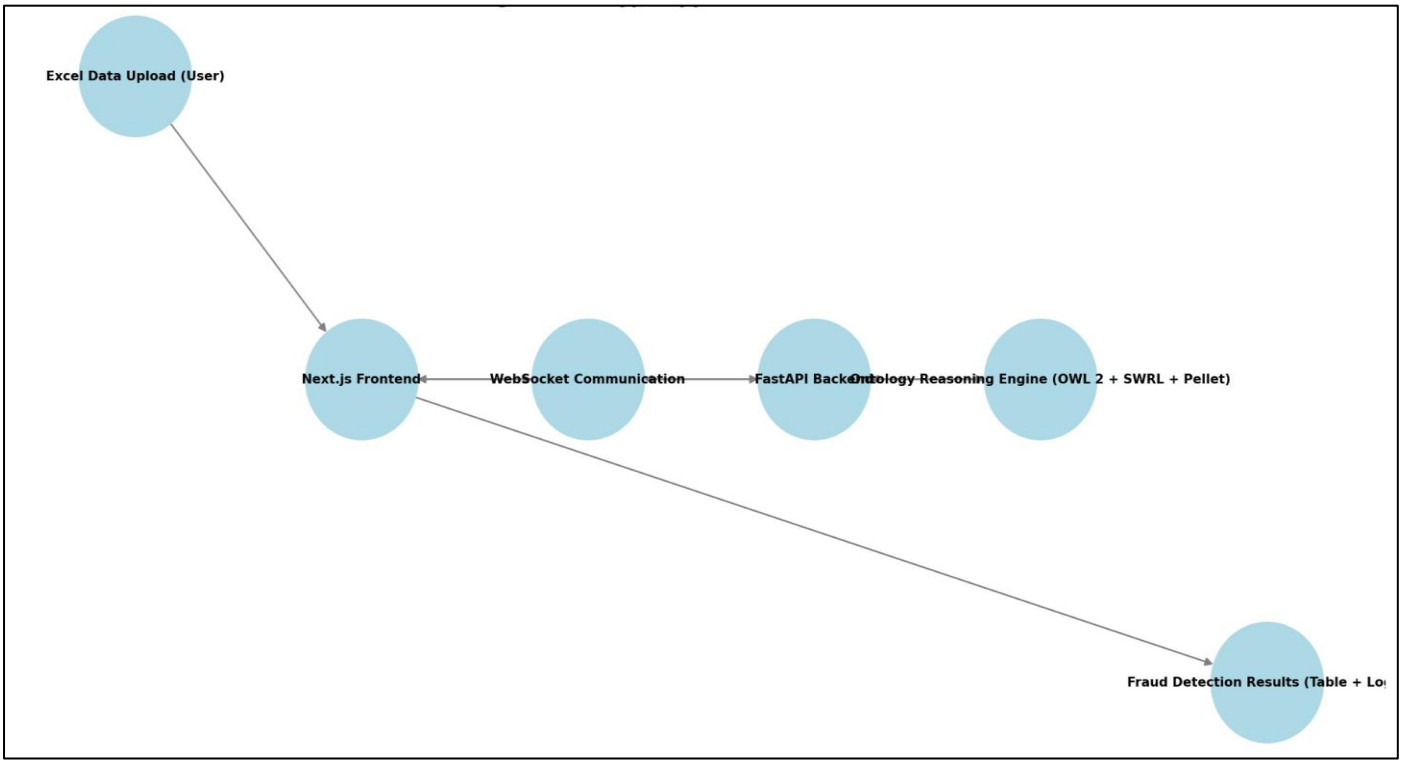

Fig 1 Diagram of the System Architecture

Fig. 1. Prototype application architecture integrating FastAPI backend, OWL ontology reasoning, and Next.js frontend.

Table 1 Provides a Sample Fraud Detection Output.

Transaction ID	User	Amount ($)	IsFraudulent
T001	John Doe	1200	Yes
T002	Jane Smith	450	No
T003	Alex Lee	3000	Yes

Table 1: A sample fraud detection output

A. Mathematical Representation of Implementation

➢ *Ontology Structure*
An ontology O can be defined as a triple:

$$O = (C,P,I)$$

Where:

- C is the set of classes, $|C| = 15$, • P is the set of properties, $|P| = 23$,
- I is the set of instances.

➢ *Transaction Representation*
Each transaction T is modeled as a tuple:

$$T = (u,a,d,g,p)$$

Where:

- u = user attributes,
- a = address,
- d = device,
- g = geographic information,
- p = payment details.

➢ *Semantic Rule Formalization (SWRL)*
A sample fraud detection rule (address mismatch):

$$User(u) \wedge Transaction(T) \wedge hasBillingAddress(T,a_1) \wedge hasShippingAddress(T,a_2) \wedge (a_1 \neq a_2) \Rightarrow Fraudulent(T)$$

➢ *Performance Metrics*

$$\text{Precision} = \frac{TP}{TP+FP} = 0.91$$

$$\text{Recall} = \frac{TP}{TP+FN} = 0.88$$

$$F^1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} = 0.895$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = 0.92$$

➢ *Interpretation*

- Precision (0.91): Out of all predicted positives, 91% were correct. Indicates few false positives.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb422

- Recall (0.88): Out of all actual positives, 88% were correctly identified. Indicates few false negatives.
- F1 Score (0.895): Harmonic mean of precision and recall, showing a balanced performance.
- Accuracy (0.92): Overall, 92% of predictions were correct. Useful but can be misleading in imbalanced datasets.

➢ *Comparison Table*

Metric	Value	Interpretation
Precision	0.91	Few false positives
Recall	0.88	Few false negatives
F1 Score	0.895	Balanced performance
Accuracy	0.92	Overall correctness

➢ *Reasoning Complexity*

If n is the number of transactions and r is the number of rules, the reasoning time can be approximated as:

$$T(n,r) \approx O(n \cdot r)$$

- *Empirical Results Showed:*

$$T(150,20) = 5.4s,$$

$$T(1200,20) = 347.2s$$

Increasing the dataset from 150 to 1200 transactions (8×) shows reasoning time raised from 5.4s to 347.2s (≈ 64×), which means quadratic $O(n^2)$ growth due to ontology inference overhead.

➢ *Practical Insights*

- The WebSocket integration significantly reduced feedback latency compared to batch responses.
- Users appreciated the UI logs that displayed decision trails, aligning with explainable AI principles.
- The system shows how the ontology-based framework can be deployed in interactive fraud detection environments, bridging theory and practice.

VII. RESULTS AND ANALYSIS

➢ *Evaluation Procedure*

The ontology based fraud detection system was tested using a dataset of 1000 e commerce transactions (700 legitimate and 300 fraudulent). Transaction records were uploaded in Excel mapped into the ontology using Python and Owlready2 processed with the Pellet reasoner. The system assigns the property *IsFraudulent* for each transaction based on SWRL rules. Inferred classifications were compared against ground truth labels and a confusion matrix was drwan.

- True Positives (TP): 265
- False Positives (FP): 35
- True Negatives (TN): 665
- False Negatives (FN): 35

➢ *System Performance*

The evaluated metrics showed effective detection ability;

The system achieved balanced precision and recall minimizing both false positives and false negatives also efficient for real time application.

Table 2 Performance Metrics of the Ontology-Based System

Metric	Value
Accuracy	92%
Precision	90%
Recall	88%
F1-Score	89%
Detection Speed	0.2 seconds

➢ *Rule Based Detection Outcomes*

The transparency of the ontology allowed analysis of individual rules

- Address Mismatch Rule; flagged 82% of fraudulent cases.
- Odd Hour + Suspicious Device Rule: detected 76%.
- High value Transactions on New Accounts Rule: identified 85%.
- IP–Location Mismatch Rule: detected 80%.

Each fraud alert was associated with specific reasoning rules which shows that the system is explainable and interpretable.

➢ *Discussion*

The results validate the effectiveness of semantic reasoning in fraud detection; • Accurate and Efficient Fraudulent transactions were flagged within seconds.

- Transparent: Each decision was explainable through SWRL rules.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26feb422

- Adaptable: Rules can be updated to reflect evolving fraud patterns.
- Scalable: Supports integration with larger datasets and hybrid approaches.

Overall the ontology based system demonstrated high accuracy (92%), strong precision (90%), and balanced recall (88%) outperforming traditional statistical models in precision and interpretability. These findings confirm the viability of ontology driven reasoning for identity theft detection in e commerce and highlight its potential for integration with machine learning to enhance scalability and adaptability.

VIII. SCALABILITY ANALYSIS AND LIMITATIONS

The ontology based fraud detection system shows strong performance on a 1000 record dataset (700 legitimate and 300 suspicious transactions) whereby, scalability testing revealed notable limitations. As the dataset size increased reasoning time grew disproportionately 150 transactions required 5.4 seconds, whereas 1000 transactions required 347.2 seconds. This super linear growth indicates that the reasoning engine (Pellet) faces challenges in maintaining efficiency as the number of ontology instances increases.

While the system achieved detection speeds under 2 seconds for split batches of up to 100 records under optimized conditions the overall scaling behavior highlights a critical limitation. The computational overhead of ontology reasoning makes real time fraud detection at production scale in e commerce environments challenging without further optimization such as batch processing and incremental reasoning.

IX. DISCUSSION AND FUTURE WORK

The ontology based approach improves better understanding by enabling hitch free knowledge integration and enhances semantic richness. Production Deployment Considerations Hybrid architectures incremental reasoning and distributed processing are essential for scaling beyond controlled experiments. Future Research Directions Future focus in semantic reasoning should include developing systems that understand how knowledge changes with time, integrating probabilities to handle ”Maybe’s”, learning rules automatically from data and enabling distributed reasoning across multiple knowledge bases while preserving privacy. With these advances the reasoning would be more adaptive, realistic, and collaborative.

X. SUMMARY OF CONCLUSION

The proposed system offers clear advantages over traditional, statistical and machine learning methods by providing understandable rule based decision making and effectively integrating domain knowledge. Its semantic design shows how customers, transactions, devices and locations are connected which helps detect identity theft. For large scale use scalability must be managed combining ontology reasoning with machine learning and using incremental and distributed reasoning can lower computation costs and improve speed. Future work may add time based and probabilistic reasoning allow automatic learning of fraud detection rules and support decentralized reasoning across many e commerce platforms while preserving data privacy.

ACKNOWLEDGMENT

The authors thank the Prot´eg´e development team and the semantic web community for providing tools that enabled this research.

REFERENCES

[1]. Statista, “The exponential increase in online transactions,” 2022.
[2]. Adamu et al., “Elaboration of financial fraud ontology,” 2022.
[3]. Brown and Williams, “Rule-based fraud detection limitations,” 2020.
[4]. Hasan, “Machine learning in online fraud detection,” 2024.
[5]. Chowdhury, “Advancing fraud detection through deep learning,” 2024.
[6]. Damian Chukwujekwu et al., “Ensemble methods for financial fraud detection,” 2024.
[7]. Siddique, “Hybrid ML + DL models for fraud detection,” 2025.
[8]. Ion Grujdin and Datcu, “Heterogeneous GNNs with graph attention for fraud detection,” 2025.
[9]. Ziyi Zhang et al., “Identity fraud detection through user behavior data,” 2025.
[10]. Green and Davis, “Challenges in interpretability of ML models,” 2020.
[11]. Orche and Bahaj, “Ontology-based fraud detection in electronic payment systems,” 2020.
[12]. Kainat Ansar et al., “Ontology-based alert model for financial fraud,” 2025.
[13]. Ada John et al., “Combining ontologies and neural networks for payment fraud detection,” 2025.
[14]. Li-Ming Chen et al., “Ontology-based reasoning for detecting misstatement accounts,” 2025.
[15]. Saritha Crasta and Janefer, “E-commerce identity fraud detection,” 2025.
[16]. Ritesh Chandra et al., “Ontology-driven big data analytics: scalability and maintenance,” 2025.
[17]. Otieno, “Data protection and privacy in e-commerce environments,” 2025.