Technical Summary Report

Key Insights from Machine Learning Applications in E-Commerce

Presented By: Mehar Sukthi Buruguru, Shyamalan Kannan Rupeshwar Rao, Nandan Varma Pericharla

Focus Areas: Customer Churn Prediction, Fraud Detection, and Recommendation Systems

1. Introduction

The rapid growth of e-commerce has intensified challenges such as customer retention, fraud mitigation, and personalized user experiences. Machine learning (ML) has emerged as a transformative tool to address these issues, offering data-driven solutions for businesses to optimize operations. This report summarizes three technical papers that explore ML applications in:

- Customer Churn Prediction using hybrid ensemble models,
- Fraud Detection through systematic literature analysis,
- Product Recommendation Systems leveraging dimensionality reduction and ML algorithms.

These studies highlight advancements in accuracy, scalability, and practical implementation, while addressing shared challenges like data imbalance and computational complexity.

2. Paper Summaries

2.1. Hybrid Ensemble-Fusion Model for Customer Churn Prediction

Source: [Nature Scientific Reports]

Objective: Improve churn prediction accuracy using ensemble techniques.
Methodology:

- Combines 17 ML algorithms (e.g., SVM, Random Forest, Neural Networks) into a hybrid Ensemble Fusion framework.
- Evaluates performance via accuracy (95.35%), AUC (91%), and F1-score (96.96%).

Key Contributions:

- Outperforms standalone models (e.g., Logistic Regression, Gradient Boosting).
- Reduces overfitting through diversity in base learners.

No	ML algorithm	Precision	Recall	Accuracy	F1-score
1	Random forests[1,2]	0.94660846	0.989123	0.942994	0.967399
2	K-nearest neighbors[5]	0.8984902	0.981404	0.889289	0.938118
3	Gradient boosting classifier [6,7]	0.95816327	0.98842	0.9532	0.96306
4	Logistic regression[3]	0.87862377	0.967719	0.858086	0.921022
5	MLPClassifier(activation = 'logistic')[1]	0.94264507	0.980351	0.932193	0.961128
6	MLPClassifier(activation = 'tanh')[1]	0.93855503	0.975439	0.924392	0.956641
7	MultinomialNB classifier[8,9]	0.86323214	0.987719	0.855686	0.921289
8	BernouiliNB classifier[8,10]	0.85508551	1	0.855086	0.921883
9	GaussianNB classifier[8,9]	0.85508551	1	0.855086	0.921883
10	DecisionTreeClassifier (CART)[3]	0.95308642	0.94807	0.915692	0.950572
11	DecisionTreeClassifier (ID3)[3]	0.95149385	0.949825	0.915692	0.950658
12	SVM classifer (Linear)[10–12]	0.85508551	1	0.855086	0.921883
13	SVM classifer (Poly)[10–12]	0.92019704	0.983158	0.912691	0.950636
14	SVM classifer (RBF)[10–12]	0.92028749	0.988421	0.916892	0.953138
15	SVM classifer(sigmoid)[10–12]	0.8604878	0.928421	0.810081	0.893165
16	Adaboost classifier[6,7]	0.94765282	0.984561	0.940294	0.965755
17	ExtraTreesClassifier[5]	0.92671706	0.989474	0.924092	0.957068
18	**Our model***	**0.960088**	0.989013	**0.953533**	**0.969631**

2.2. E-Commerce Fraud Detection: A Systematic Review

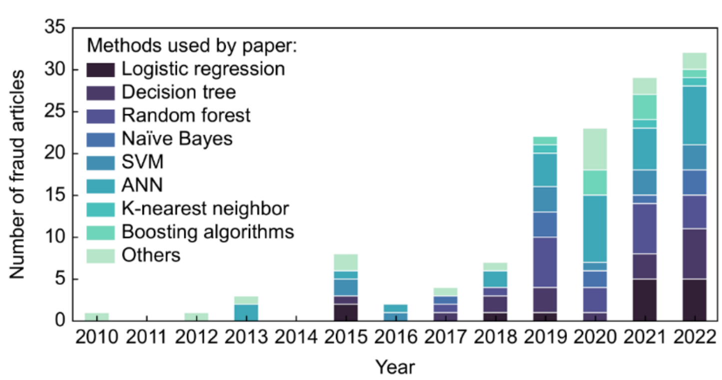
Source: [IEEE Xplore]
*Objective***:** Identify trends and gaps in ML-driven fraud detection.
*Methodology***:**

- PRISMA framework applied to analyse 101 studies (2013–2023).
- Highlights Artificial Neural Networks (ANNs) as a dominant approach.

Key Findings:

- Data scarcity and class imbalance limit model generalizability.
- Platform-specific fraud patterns (e.g., eBay vs. Amazon) require tailored solutions.

Visualization:



2.3. Product Recommendation System Using PCA and ML

Source: [IJCA]

Objective: Enhance recommendation accuracy with feature reduction.

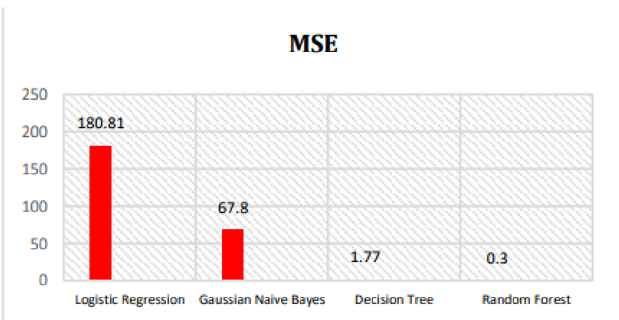
Methodology:

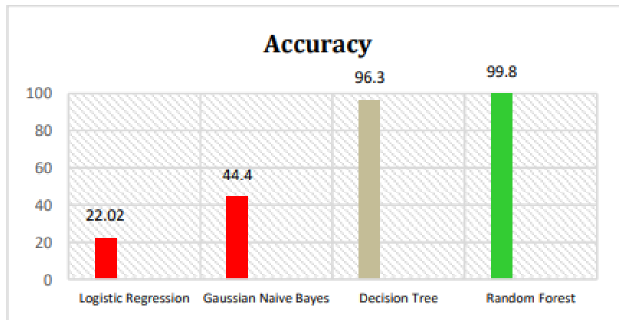
- Applies Principal Component Analysis (PCA) to reduce dimensionality.
- Tests four ML models; Random Forest (RF) achieves 99.6% accuracy.

Key Contributions:

- RF outperforms Gaussian Naive Bayes and Logistic Regression.
- PCA minimizes computational costs without sacrificing performance.
- The study demonstrates that applying Principal Component Analysis (PCA) for dimensionality reduction in product recommendation systems effectively minimizes computational costs without sacrificing performance.

Visualization:





3. Comparative Analysis

3.1. Key Techniques and Performance

Aspect	**Churn Prediction**	**Fraud detection**	**Recommendation Systems**
ML Algorithms	Hybrid Ensembles	ANNs, Random Forests	Random Forest, PCA
Top Accuracy	95.35%	Varies (ANN~90%)	99.6% (Random Forest)
Data Challenges	Class Imbalance	Data Scarcity, Imbalance	High Dimensionality
Preprocessing	Resampling, Normalization	Synthetic data generation	Feature Scaling, PCA

3.2. Common Themes and Challenges

Data Quality: All studies emphasize the need for balanced datasets.

Solutions: Synthetic data (fraud detection), resampling (churn prediction).

Model Complexity:

- Ensemble methods improve robustness but increase computational costs.
- PCA mitigates complexity in recommendation systems.

Real-World Scalability:

- Fraud detection requires real-time processing; churn prediction needs proactive alerts.

4. Conclusion

The three papers demonstrate ML's pivotal role in solving critical e-commerce challenges:

- **Hybrid Ensemble-Fusion Models** achieving superior accuracy in churn prediction, outperforming standalone classifiers with a **95.35% accuracy and 96.96% F1-score** [7].
- **Systematic Reviews on Fraud Detection**, emphasizing the dominance of **Artificial Neural Networks (ANNs)** while addressing challenges like class imbalance and data scarcity [16].
- **PCA-Enhanced Recommendation Systems**, where **Random Forest models with PCA** achieved **99.6% accuracy**, optimizing performance while reducing computational costs [5].

Future Directions:

- **Integration of Deep Learning for Real-Time Fraud Detection** – ANN-based models are increasingly utilized for fraud detection, offering high accuracy and adaptability in identifying fraudulent transactions in real-time
- **Personalized AI-driven Recommendations** – Combine reinforcement learning with customer behaviour analysis for dynamic and adaptive product suggestions.
- **Standardized Datasets and Evaluation Metrics** – Establish industry-wide benchmarks for fraud detection models to improve comparability and reliability.
- **Bias and Fairness in AI** – Implement techniques to detect and mitigate bias in fraud detection and credit scoring models to ensure fair decision-making.