CASE STUDY

IMPLEMENTING AN INTRUSION PREVENTION SYSTEM (IPS) TO PROTECT A HIGH-TRAFFIC E-COMMERCE PLATFORM

Presented by

Md Hashir Imteyaz	:2141009001
Rohit Raj	:2141018031
Sulagnna Mohanty	:2141013223
Swati Nanda	:2141014015

Presentation Outline

- Introduction
    - Overview
    - Motivation
- Literature Survey
- Background Analysis
- Proposed Solution
- Result and Analysis
- Conclusion

Introduction

Overview:

- E-commerce platforms are prime targets for cyber-attacks due to valuable transactions and sensitive information.
- Ensuring e-commerce security is vital for customer trust, business continuity, and financial data protection.
- High-traffic e-commerce platforms need scalable and secure solutions to handle enormous data volumes.

Introduction

Motivation:

- **Enhance Security Measures:** Improve upon traditional security measures that are inadequate for handling the complexity and volume of modern e-commerce transactions.
- **Reduce Financial and Reputational Risks:** Minimize the financial loss, reputational damage, and legal consequences businesses face due to data breaches.
- **Ensure Scalability:** Develop a scalable security solution capable of handling the high data volumes generated by high-traffic e-commerce platforms without performance degradation.

Existing Research Work

AUTHOR	TECHNIQUE	RESULT
Dinesh Chowdary Attota et al.[1]	Multi-View Federated Learning-based ID(MV-FLID)	94.17% accuracy
Larriva-Novo et al. [2]	Preprocessing with content characterization multi-layer perceptron for detection	KDD99 with 95.5% accuracy
Dhanke JyotiAtul et al [3]	NB,MLP,Multi nomial logistic regression	MLP with 92.66% accuracy
Maonan Wang et al [4]	SHapely Additive exPlanations(SHAP)	83.0% accuracy

Problem Identification

TABLE 2 Limitations of Traditional Security Measures

TOPIC	KEY POINT	SOURCE
Signature-Based Detection Systems	• Effective for known threats • Fails against new threats • High maintenance due to constant updates	• Axelsson, S. (2000) • Garcia-Teodoro, P., et al. (2009)

- **Effective Ensemble Learning Techniques**:Researchers have found that Ensemble Learning (EL) methods, such as Random Forest (RF), often provide superior performance compared to other models like Logistic Regression (LR) and Support Vector Machines (SVM)

- **Importance of Comprehensive Attributes**:Including a diverse set of attributes helps in creating a balanced dataset that represents all aspects of network traffic

- **Data Quality Issues**:Thorough data cleaning like missing values, removal of duplicate entries to prevent overfitting , filter out noise and irrelevant data

Objective

- Implement an IPS that provides real-time monitoring and automated response to detect and block various types of cyber threats, including DDoS attacks, SQL injection, and other vulnerabilities, to safeguard the integrity and security of the e-commerce platform.

- Implement an Intrusion Prevention System (IPS) that utilizes Machine Learning for advanced threat detection and integrates with a Firewall for robust protection, ensuring the system can handle high traffic volumes efficiently while maintaining optimal performance and a seamless user experience.

Proposed Architecture & Implementation plan

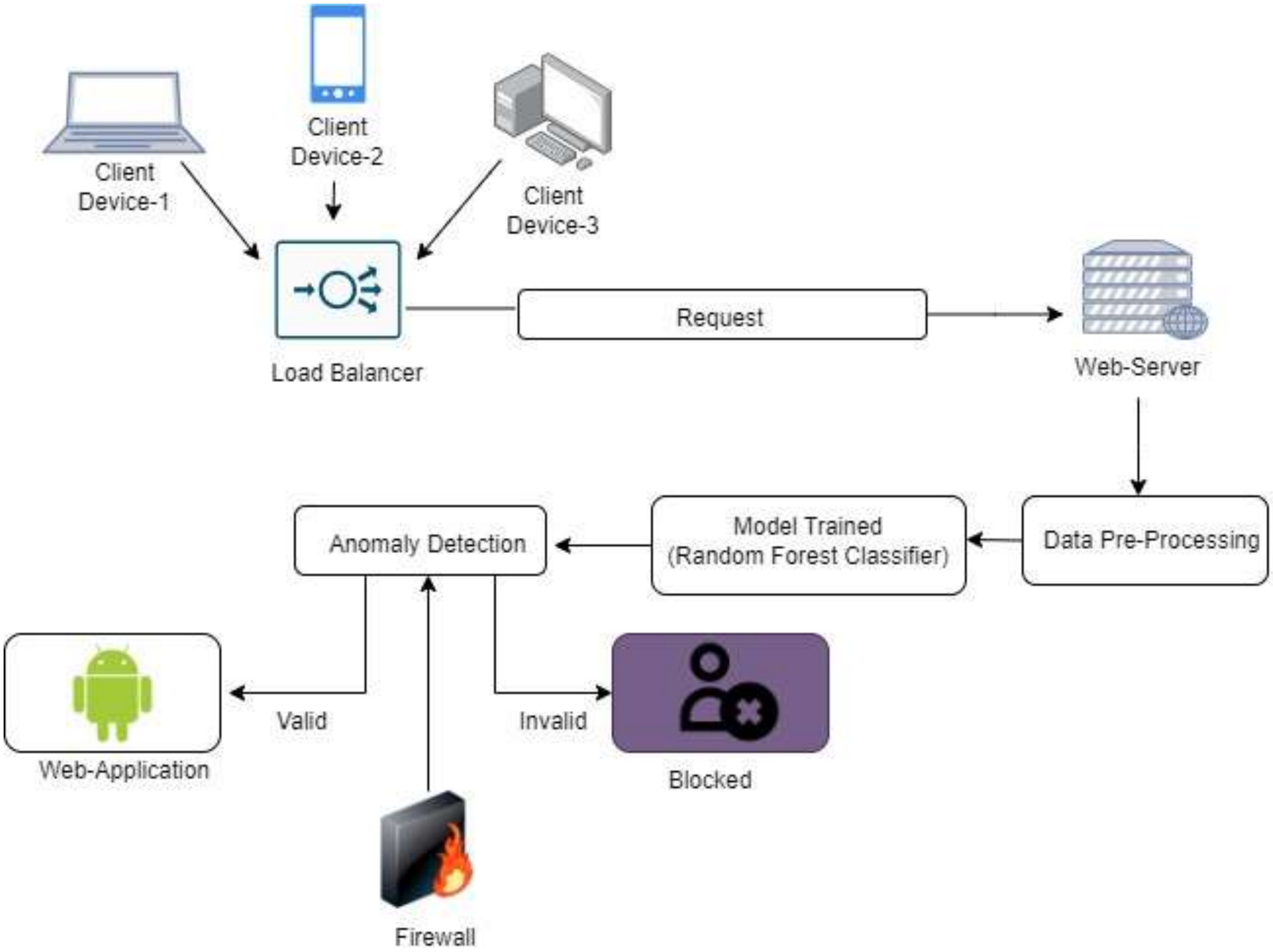


FIG 1 Architecture for IPS in E-Commerce website for high traffic

Machine Learning Workflow

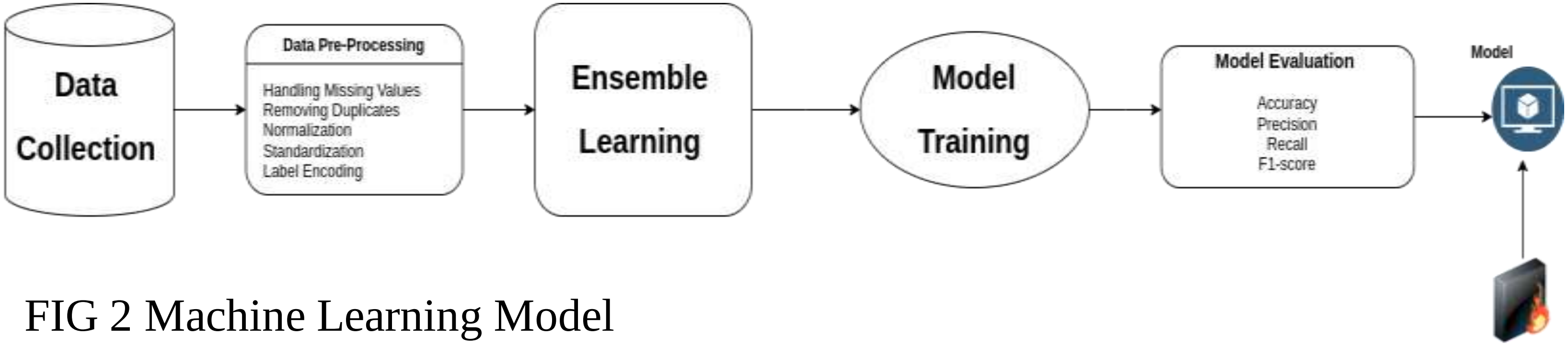


FIG 2 Machine Learning Model

- **Data Collection**:The dataset has been collected form Kaggle. It consist the information of Network and Traffic Details, Traffic Analysis, Session and Response Metrics, User and Request Information and Anomaly Detection. The dataset has 42 columns with 54% normal instance and 46% attacks
- **Data preprocessing**:We check for missing values and drop unnecessary columns then used LabelEncoder to transform categorical variables into numerical values. We also used train_test_split to create a 70-30 split of dataset

Machine Learning Workflow

- **Ensemble learning:** It is a machine learning technique that aggregates two or more learners (e.g. random forest,regression models, neural networks) in order to produce better predictions.
- **Model Training**: This proposed model works with five different kinds of model training method that is Logistic Regression, k-Nearest Neighbors (k-NN),Decision Tree, Support Vector Machine (SVM) and Random Forest
- **Model Evaluation**: The model uses five performance parameters to analyze the performance i.e. Training Accuracy, Test Accuracy, F1 Score , Precision and Recall.

Intrusion Prevention

- Firewall and Pickle
  - Used Python's pickle module to save and load a Random Forest Classifier model. This model is then integrated into a firewall system to identify and block malicious data.
  - Essentially, the classifier predicts whether data is harmful or not, and the firewall uses these predictions to prevent malicious data from entering your network.
- Ensemble Learning
  - Combines multiple learning algorithms for better predictive performance.
  - Random Forest, a popular ensemble method, constructs multiple decision trees and aggregates their results.

Dataset Description and Analysis

Dataset Description:

- **Network and Traffic Details:** Includes information about timestamps, source and destination IP addresses, source and destination ports, protocols used (e.g., HTTP, HTTPS), packet sizes, and specific TCP/IP flags.
- **Traffic Analysis**: Features such as traffic type (normal, suspicious, malicious), request methods (GET, POST), and the number of requests made by a source IP.
- **Session and Response Metrics**: Contains session duration, response times, and HTTP status codes.
- **User and Request Information**: Includes user-agent strings and payload data, detailing the client software and the actual data transferred.
- **Anomaly Detection**: Anomaly scores indicating the likelihood of traffic being anomalous or malicious, possibly generated by preliminary detection algorithms.

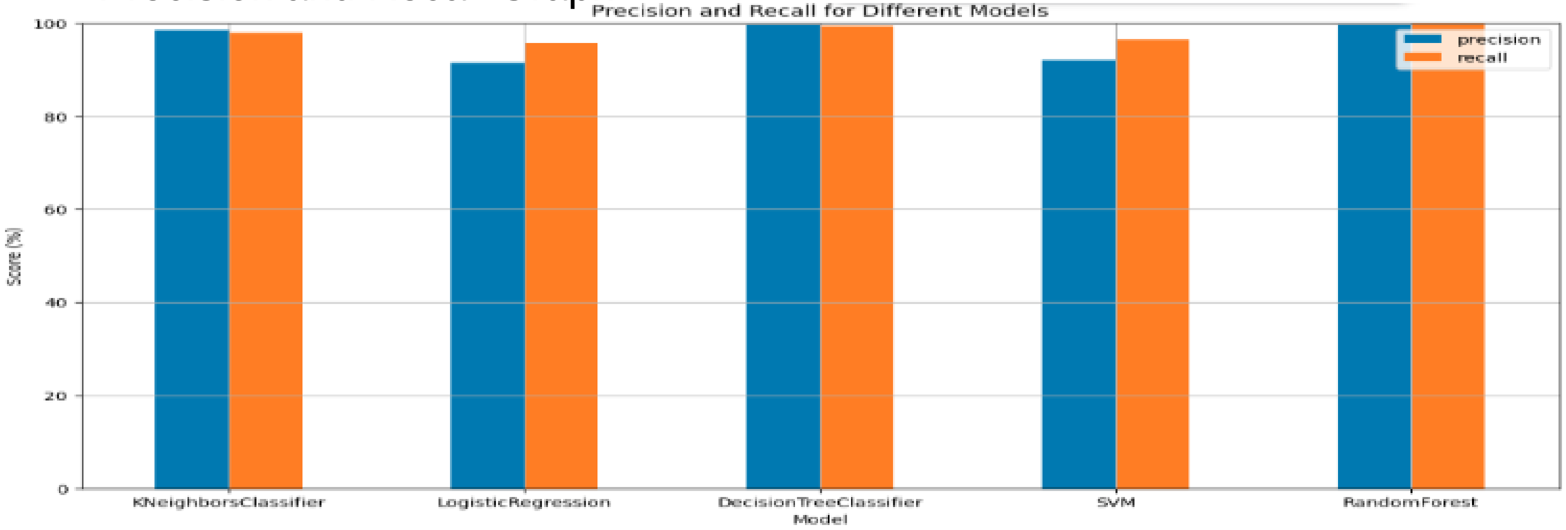
Dataset Description and Analysis

Analysis:

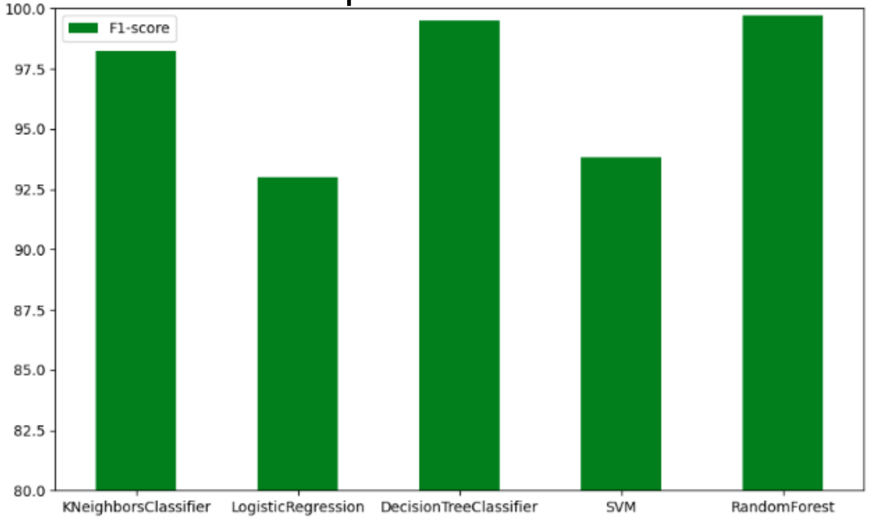
- Ensemble learning improves model accuracy by combining multiple models.
- Random Forest, which uses multiple decision trees, is a common ensemble method.
- The Random Forest classifier is used for traffic detection and prevention.
- Another ensemble model, such as stacking classifier was also implemented.
- However, stacking took more time to evaluate on the same data.
- Overall, while the Random Forest classifier is accurate, it is less time-consuming compared to stacking ensemble as well as boosting ensemble.

Results and Comparison

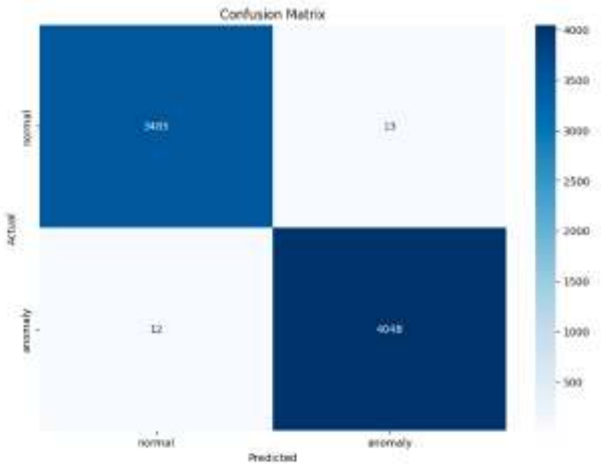
Precision and Recall Graph



F1 Score Graph



Confusion Matrix



Results and Comparison

TABLE 3 Comparing results of different ML Models

Model	Accuracy	F1 Score	Precision	Recall
KNeighborsClassifier	98.24%	98.10%	98.48%	97.91%
Logistic Regression	92.31%	93.50%	91.35%	95.72%
DecisionTree Classifier	99.38%	99.49%	99.5%	99.48%
SVM	93.56%	94.11%	91.95%	96.34%
Random Forest	99.64%	99.68%	99.53%	99.83%

Output after implementation of IPS

```
Model saved successfully.
Model loaded successfully.
Malicious data detected at indices:[0, 1, 3, 4, 12, 13, 19, 20, 21, 24 ...]
Firewalls blocked those data
```

Conclusion & Future Scope

- The architecture enhances security for high-traffic e-commerce platforms through advanced feature engineering, comprehensive model training, smart preprocessing, and seamless integration, with a focus on real-time threat detection and prevention.
- Random Forest offers a practical balance of accuracy and efficiency for real-time detection and prevention of traffic anomalies, outperforming the more time-consuming stacking method. It ensures robust security with high accuracy and minimal false positives.

References

[1] D. C. Attota, V. Mothukuri, R. M. Parizi, and S. Pouriyeh, An ensemble multi-view federated learning intrusion detection for IoT, IEEE Access, vol. 9, pp. 117734117745, 2021.

[2] X. Larriva-Novo, V. A. Villagrá, M. Vega-Barbas, D. Rivera, and M. S. Rodrigo, An IoT-focused intrusion detection system approach based on preprocessing characterization for cybersecurity datasets, Sen sors, vol. 21, no. 2, p. 656, Jan. 2021.

[3] D. J. Atul, R. Kamalraj, G. Ramesh, K. S. Sankaran, S. Sharma, and S. Khasim, A machine learning based IoT for providing an intrusion detection system for security, Microprocessors Microsystems, vol. 82, Apr. 2021, Art. no. 103741.

[4] M. Wang, K. Zheng, Y. Yang, and X. Wang, “An explainable machine learning framework for intrusion detection systems”, IEEE Access,vol.8, pp. 73127-73141, 2020.