iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84280

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

Cloud-Based Security Analytics SIEM Using ELK Stack and Machine Learning

Gugulothu Sai Charan[1], Dr. P Sammulal[2]
[1]*Post Graduate Student, M. Tech (Computer Networks and Information Security), Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India*
[2]*Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India*

Abstract: ***The increasing frequency and sophistication of cyberattacks have created significant challenges for organizations in protecting their digital infrastructure. Traditional Security Information and Event Management (SIEM) systems primarily depend on predefined rules and signature-based detection techniques, which are often inadequate for identifying emerging or previously unseen attack patterns. To address these limitations, this paper presents a Cloud-Based Security Analytics SIEM Using ELK Stack and Machine Learning for intelligent cyberattack detection and real-time security monitoring. The proposed framework integrates Elasticsearch, Logstash, and Kibana (ELK Stack) with supervised machine learning algorithms to automate log collection, processing, threat classification, and security visualization. The CICIDS2017 dataset, containing approximately 2.5 million network traffic records, 70 selected features, and 15 attack categories, was used to train and evaluate the intrusion detection models. Random Forest, LightGBM, and XGBoost classifiers were developed and compared using multiple performance metrics. Experimental results indicate that the XGBoost model achieved the best overall performance with an accuracy of 99.86% and a Macro F1-score of 89.57%, demonstrating its effectiveness in multiclass cyberattack classification. The trained model was integrated with the ELK-based analytics platform to support automated attack prediction and centralized visualization through Kibana dashboards. The proposed system provides an efficient, scalable, and intelligent approach for enhancing cybersecurity monitoring by combining cloud-based log analytics with machine learning-driven threat detection. The framework can assist Security Operations Centers (SOCs) in improving incident detection, accelerating security analysis, and supporting proactive cyber defense.***

Keywords: Security Information and Event Management (SIEM), ELK Stack, Machine Learning, XGBoost, Random Forest, LightGBM, Cybersecurity, Intrusion Detection System, CICIDS2017, Elasticsearch, Logstash, Kibana.

I. INTRODUCTION

The rapid advancement of digital technologies, cloud computing, and enterprise networking has significantly increased the volume and complexity of cybersecurity threats. Organizations across government, healthcare, education, finance, and industrial sectors generate massive amounts of security logs from firewalls, intrusion detection systems, servers, endpoints, and network devices. These logs contain valuable information for identifying malicious activities; however, manually analyzing such large-scale data is impractical and time-consuming. As cyberattacks continue to evolve in sophistication, organizations require intelligent Security Information and Event Management (SIEM) solutions capable of providing real-time threat detection, centralized log management, and actionable security insights.

Traditional SIEM platforms primarily rely on predefined rules, signatures, and correlation mechanisms to detect security incidents. Although these approaches are effective against known threats, they often fail to identify previously unseen attacks, zero-day exploits, and complex attack patterns that do not match existing rules. Furthermore, maintaining rule-based detection systems requires continuous updates, increasing operational complexity and resulting in higher false-positive rates. These limitations highlight the need for intelligent SIEM systems that can automatically learn attack characteristics from security data and improve detection accuracy through data-driven techniques.

Recent advances in Machine Learning (ML) have provided effective approaches for enhancing intrusion detection by learning complex relationships within network traffic and security events. Supervised learning algorithms such as Random Forest, XGBoost, and LightGBM have demonstrated excellent performance in identifying multiple categories of cyberattacks while maintaining high classification accuracy. When integrated with SIEM platforms, these models enable automated threat classification, faster incident response, and improved security analytics.

The ELK Stack, comprising Elasticsearch, Logstash, and Kibana, has become one of the most widely adopted open-source platforms for centralized log collection, storage, search, and visualization. Logstash ingests and processes security logs from multiple sources, Elasticsearch provides scalable indexing and high-performance search capabilities, and Kibana offers interactive dashboards for monitoring security events. The integration of machine learning with the ELK Stack enables the development of intelligent SIEM platforms capable of combining real-time log analytics with predictive attack detection.

This paper presents a Cloud-Based Security Analytics SIEM Using ELK Stack and Machine Learning, an intelligent framework designed to improve cyberattack detection through the integration of centralized log management and supervised machine learning techniques. The proposed system utilizes the CICIDS2017 benchmark dataset, which contains approximately 2.5 million network traffic records, 70 selected network flow features, and 15 attack categories. Three machine learning algorithms—Random Forest, XGBoost, and LightGBM—were trained and evaluated for multiclass intrusion detection. Among the evaluated models, XGBoost achieved the highest performance with an accuracy of 99.86% and a Macro F1-score of 89.57%. The trained model was integrated with the ELK-based analytics platform to support automated attack prediction and real-time visualization through Kibana dashboards, providing a scalable and efficient solution for modern Security Operations Centers (SOCs). The experimental results demonstrate that the proposed framework enhances threat detection capabilities while simplifying security monitoring through centralized analytics and intelligent decision support.

A. Motivation of the study

The increasing frequency of ransomware attacks, distributed denial-of-service (DDoS) attacks, phishing campaigns, and advanced persistent threats has made cybersecurity one of the most critical concerns for modern organizations. Security teams are required to monitor large volumes of heterogeneous log data generated by diverse network devices and applications. Traditional SIEM systems are effective for monitoring known attack signatures but often struggle to detect sophisticated or previously unseen threats without extensive manual rule development.

Machine learning offers the capability to automatically identify hidden attack patterns by learning from historical network traffic data, thereby reducing dependence on manually defined rules. At the same time, the ELK Stack provides an efficient platform for centralized log management and visualization, enabling security analysts to investigate incidents more effectively. These developments motivated the design of an integrated cloud-based SIEM framework that combines the strengths of ELK Stack and machine learning to improve threat detection accuracy, automate security analytics, and support real-time monitoring through interactive dashboards.

B. Objectives

The primary objective of this research is to develop an intelligent cloud-based Security Information and Event Management (SIEM) framework that integrates the ELK Stack with supervised machine learning techniques for efficient cyberattack detection and security analytics. The study aims to design a centralized log management system capable of collecting, processing, storing, and visualizing security events from multiple sources while improving attack detection through data-driven classification models.

Specifically, the objectives of this work are to preprocess the CICIDS2017 dataset for intrusion detection, train and evaluate Random Forest, XGBoost, and LightGBM classifiers for multiclass attack classification, compare their performance using standard evaluation metrics, integrate the best-performing model into the ELK-based analytics pipeline, and provide real-time visualization of security events through Kibana dashboards. The proposed framework is intended to enhance threat detection accuracy, simplify security monitoring, and support faster incident analysis in Security Operations Centers (SOCs).

C. Problem Statement

Modern organizations generate enormous volumes of security logs from firewalls, servers, endpoints, network devices, and cloud applications. Although conventional Security Information and Event Management (SIEM) platforms centralize these logs, their detection capabilities are primarily based on predefined rules and signature matching. Such approaches are effective for identifying known threats but often fail to detect new, evolving, or sophisticated cyberattacks. As the scale and complexity of network environments continue to increase, manually maintaining detection rules becomes challenging, resulting in delayed incident response and a higher likelihood of false positives or missed attacks.

Machine learning has demonstrated significant potential for improving intrusion detection by learning complex attack patterns directly from network traffic data. However, many existing research solutions focus only on offline model development and do not integrate machine learning with operational SIEM platforms for centralized security monitoring and visualization.

Therefore, there is a need for a scalable framework that combines real-time log management, intelligent attack classification, and interactive security analytics within a unified platform.

To address these challenges, this research proposes a **Cloud-Based Security Analytics SIEM Using ELK Stack and Machine Learning**, which integrates the ELK Stack with supervised machine learning models to automate log analysis, classify multiple cyberattack categories, and present security insights through centralized Kibana dashboards. The proposed framework aims to improve cyberattack detection accuracy while supporting efficient security monitoring and incident investigation in modern enterprise environments.

II. LITERATURE REVIEW

A. Traditional Security Information and Event Management Systems

Security Information and Event Management (SIEM) systems have become an essential component of modern cybersecurity by providing centralized log collection, event correlation, security monitoring, and incident response. Commercial SIEM platforms such as IBM QRadar, Splunk Enterprise Security, ArcSight, and Microsoft Sentinel enable organizations to collect security events from diverse sources and analyze them through predefined correlation rules. These platforms improve visibility into network activities and assist security analysts in identifying malicious behavior.

Despite their widespread adoption, conventional SIEM systems primarily depend on rule-based detection and signature matching techniques. Such approaches are highly effective for identifying previously known attack patterns but often struggle to detect zero-day attacks, advanced persistent threats (APTs), and rapidly evolving cyber threats. In addition, maintaining large rule sets requires continuous manual updates, increasing operational complexity and generating a significant number of false-positive alerts. These limitations have encouraged researchers to investigate intelligent detection mechanisms based on machine learning and behavioral analytics.

B. ELK Stack-Based Security Analytics

The ELK Stack, consisting of Elasticsearch, Logstash, and Kibana, has emerged as a powerful open-source platform for centralized log management and security analytics. Elasticsearch provides distributed indexing and high-performance search capabilities, Logstash supports data ingestion and preprocessing from multiple log sources, and Kibana offers interactive dashboards for real-time visualization and analysis of security events. Due to its scalability and flexibility, the ELK Stack has been widely adopted by organizations for security monitoring and operational analytics.

Several research studies have demonstrated the effectiveness of the ELK Stack in collecting and visualizing security events from enterprise networks. However, many implementations rely solely on log aggregation and visualization without incorporating intelligent attack detection techniques. Consequently, security analysts must manually interpret dashboards and investigate suspicious activities, limiting the system's ability to automatically identify sophisticated cyber threats. Integrating machine learning models with the ELK Stack can significantly enhance detection capabilities by enabling automated classification of malicious network traffic and supporting proactive security monitoring.

C. Machine Learning for Intrusion Detection

Machine learning has become an effective approach for intrusion detection because of its ability to identify complex relationships within network traffic and distinguish malicious activities from legitimate communication. Supervised learning algorithms, including Decision Trees, Support Vector Machines, Random Forest, XGBoost, and LightGBM, have been widely investigated for network intrusion detection using benchmark datasets such as KDD Cup 1999, NSL-KDD, UNSW-NB15, and CICIDS2017.

Among these algorithms, ensemble learning methods have demonstrated superior classification performance due to their capability to combine multiple decision trees and reduce prediction errors. XGBoost and LightGBM have achieved particularly strong results in multiclass intrusion detection by efficiently handling high-dimensional network traffic features and class imbalance. Nevertheless, many published studies focus primarily on offline model evaluation without integrating trained models into operational SIEM environments. As a result, their practical applicability for real-time security monitoring remains limited.

D. Cloud-Based Security Analytics

Cloud computing has transformed enterprise information technology by providing scalable infrastructure, distributed storage, and on-demand computing resources. As organizations increasingly migrate critical services to cloud environments, the need for scalable security monitoring solutions has grown significantly.

Cloud-based SIEM platforms enable centralized log collection from geographically distributed systems while supporting continuous monitoring and rapid incident investigation.

Recent research has explored cloud-native security analytics by combining centralized log management with artificial intelligence to improve threat detection. Although these approaches demonstrate improved scalability, many existing frameworks emphasize cloud infrastructure monitoring rather than integrating machine learning-based attack classification with open-source SIEM technologies. Furthermore, several proposed solutions depend on proprietary commercial platforms, limiting accessibility for academic institutions and small organizations. These challenges highlight the need for cost-effective and intelligent cloud-based security analytics frameworks using open-source technologies.

E. Research Gap

The review of existing literature indicates that traditional SIEM systems provide effective log management and event correlation but remain largely dependent on predefined rules for threat detection. Similarly, many machine learning-based intrusion detection studies demonstrate excellent classification accuracy under laboratory conditions but do not integrate their models into operational SIEM platforms capable of supporting real-time security monitoring. Although the ELK Stack offers powerful log management and visualization capabilities, many existing implementations focus primarily on centralized monitoring without incorporating automated attack classification.

To address these limitations, this research proposes a Cloud-Based Security Analytics SIEM Using ELK Stack and Machine Learning, integrating centralized log collection, scalable data processing, supervised machine learning, automated attack prediction, and interactive Kibana dashboards within a unified framework. The proposed approach combines the operational strengths of the ELK Stack with intelligent multiclass intrusion detection using the CICIDS2017 dataset, thereby providing an efficient and scalable solution for modern Security Operations Centers (SOCs).

III. METHODOLOGY

The proposed Cloud-Based Security Analytics SIEM Using ELK Stack and Machine Learning is designed to provide an intelligent and scalable security monitoring framework capable of detecting cyberattacks through centralized log management and supervised machine learning. The framework integrates the ELK Stack (Elasticsearch, Logstash, and Kibana) with machine learning models to automate security log collection, preprocessing, attack classification, and visualization within a unified platform.

The overall workflow begins with the collection of network traffic and security logs from multiple sources. Logstash receives and processes the incoming data by parsing, filtering, and transforming it into a structured format suitable for storage. The processed logs are indexed in Elasticsearch, which provides efficient storage and high-speed search capabilities for security events. Network traffic data derived from the CICIDS2017 dataset is preprocessed to remove inconsistencies, normalize feature values, and prepare the data for model training. Three supervised machine learning algorithms—Random Forest, XGBoost, and LightGBM—are trained and evaluated using the processed dataset. Based on experimental evaluation, the best-performing model is integrated into the prediction pipeline to classify network traffic into normal or attack categories. The prediction results are indexed back into Elasticsearch and visualized through Kibana dashboards, enabling security analysts to monitor network activities, identify suspicious events, and support rapid incident investigation.

The proposed framework combines cloud-based log analytics with intelligent cyberattack detection, providing centralized monitoring, automated threat classification, and interactive visualization to improve the efficiency of modern Security Operations Centers (SOCs).

A. Overall System Architecture

The proposed architecture consists of five major components that operate sequentially to provide real-time security monitoring and intelligent attack detection.

1) Data Collection Layer: The framework collects security logs and network traffic from multiple sources, including servers, network devices, operating systems, and simulated intrusion datasets. These logs represent different categories of security events that require centralized monitoring and analysis.

2) Log Processing Layer: Logstash acts as the primary data processing engine by collecting incoming logs, parsing event fields, filtering unnecessary information, and transforming the records into a standardized structure. This preprocessing stage ensures that security events are stored in a consistent format suitable for indexing and analysis. The Logstash pipeline supports multiple inputs such as Beats, Syslog, TCP, UDP, and HTTP, allowing integration with diverse log sources.

3) Data Storage Layer: The processed events are indexed into Elasticsearch, which provides distributed storage, rapid search functionality, and efficient retrieval of security information. Elasticsearch enables centralized management of large-scale security logs while supporting near real-time querying for security analytics.

4) Machine Learning Layer: The machine learning module processes the CICIDS2017 dataset through data preprocessing, feature selection, model training, and performance evaluation. Three supervised classification algorithms—Random Forest, XGBoost, and LightGBM—are developed and compared for multiclass intrusion detection. Feature scaling is performed using the StandardScaler technique, followed by stratified train-test splitting to preserve class distribution during model evaluation. The training pipeline stores the scaler, trained models, feature definitions, and evaluation summaries for later inference.

5) Visualization Layer: Prediction results generated by the trained model are stored in Elasticsearch and visualized using Kibana dashboards. The dashboards provide centralized monitoring of attack categories, event trends, and security analytics, enabling analysts to quickly identify suspicious activities and support incident response.

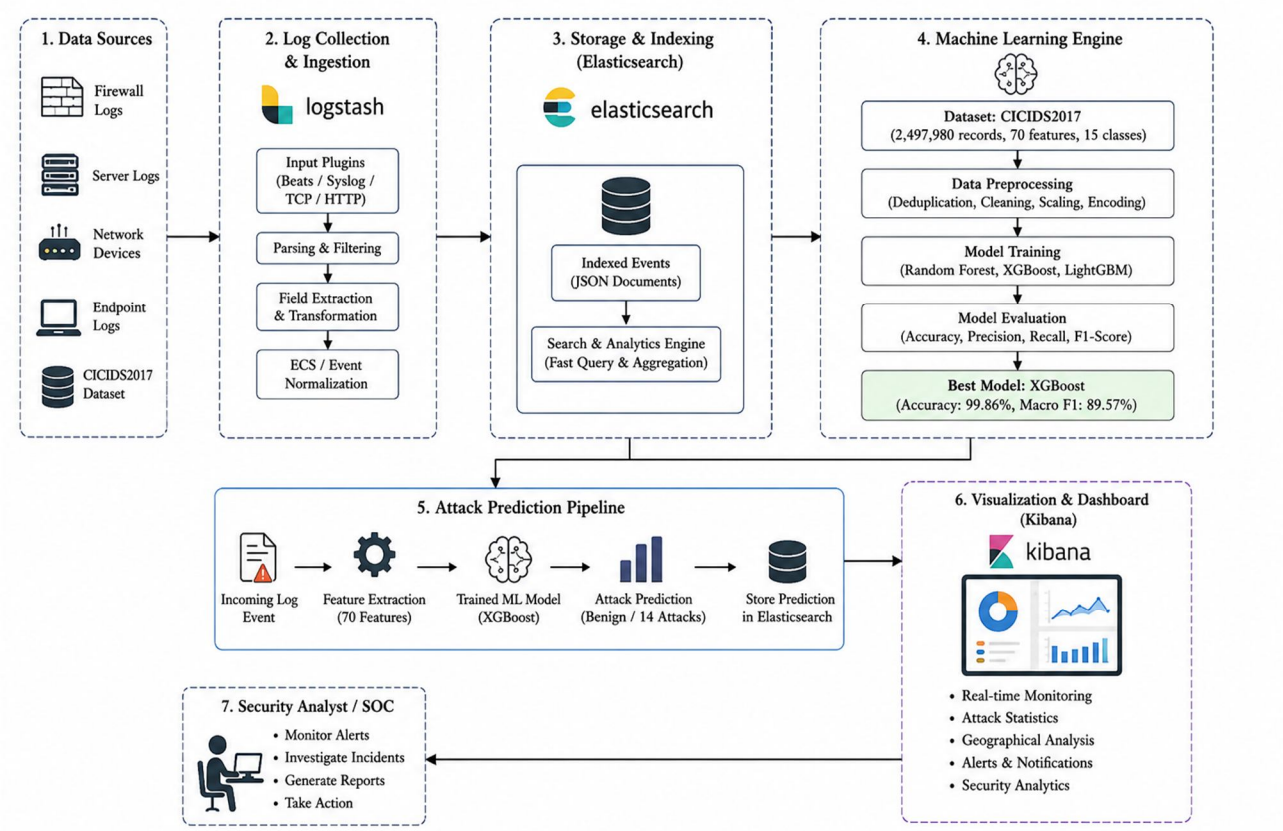


Figure 1. Overall Architecture of the Proposed Cloud-Based Security Analytics SIEM Framework Using ELK Stack and Machine Learning

B. Dataset Description and Preprocessing

The proposed Security Information and Event Management (SIEM) framework utilizes the publicly available CICIDS2017 dataset, developed by the Canadian Institute for Cybersecurity (CIC). This dataset is widely recognized as a benchmark for network intrusion detection research because it closely simulates real-world enterprise network traffic under both normal and malicious conditions. The dataset contains network flow records generated over five consecutive working days and includes a diverse range of modern cyberattacks along with legitimate user activities.

After preprocessing, the dataset contains 2,497,980 network flow records with 70 numerical features and 15 traffic classes, consisting of one benign class and fourteen attack categories, including DDoS, DoS Hulk, DoS GoldenEye, DoS Slowloris, DoS Slowhttptest, PortScan, Bot, FTP-Patator, SSH-Patator, Heartbleed, Infiltration, Web Attack–Brute Force, Web Attack–SQL Injection, and Web Attack–XSS. The dataset exhibits significant class imbalance, with benign traffic accounting for approximately 82.96% of the total records.

To improve the quality of the dataset, several preprocessing operations were performed before model training. Duplicate flow records were removed, reducing redundancy in the dataset. Missing values were eliminated by discarding incomplete records, while infinite values were replaced with valid numerical representations where necessary. Eight constant-value attributes that provided no discriminative information were removed, resulting in a final feature set of 70 informative attributes. All numerical features were converted to float32 format to reduce memory consumption during training without affecting classification performance. Finally, categorical attack labels were encoded into numerical values to facilitate supervised machine learning.

The processed dataset was divided into 80% training data and 20% testing data using a stratified sampling strategy to preserve the original class distribution. Feature normalization was performed using the StandardScaler algorithm, which was fitted only on the training dataset and subsequently applied to the testing dataset to prevent data leakage during model evaluation. This preprocessing pipeline ensured data consistency, reduced computational overhead, and improved the reliability of the machine learning models used in the proposed SIEM framework.

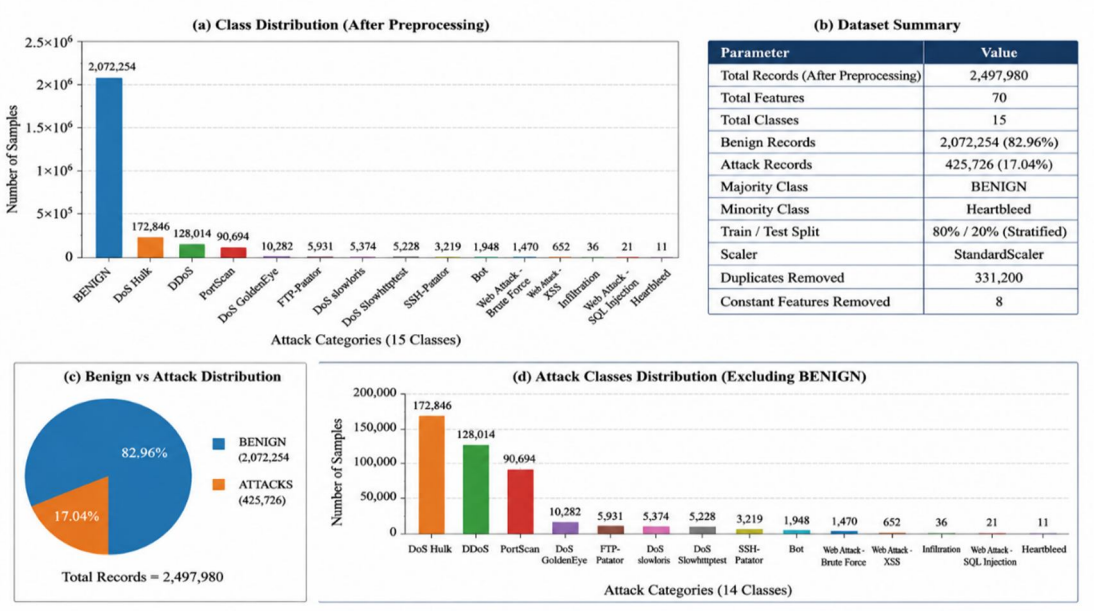


Figure 2. Distribution of the CICIDS2017 Dataset after Preprocessing

C. ELK Stack-Based Log Processing

The proposed Security Information and Event Management (SIEM) framework employs the **ELK Stack**, comprising Elasticsearch, Logstash, and Kibana, to provide centralized log collection, processing, storage, and visualization. This open-source platform enables efficient management of large volumes of security events generated from multiple network and system components.

Logstash serves as the data ingestion and preprocessing engine of the framework. It receives security logs from multiple input sources, including Beats, Syslog, TCP, and HTTP protocols. During the ingestion process, Logstash parses raw log messages, extracts relevant event attributes using filtering plugins, and transforms them into a structured format based on the Elastic Common Schema (ECS). This preprocessing stage ensures consistency across heterogeneous log sources and prepares the data for efficient indexing and analysis.

After preprocessing, the structured events are forwarded to Elasticsearch, where they are indexed and stored as searchable documents. Elasticsearch provides distributed storage, high-speed indexing, and real-time search capabilities, enabling rapid retrieval and analysis of security events. Daily indices are created to efficiently manage continuously generated logs while supporting scalable security analytics. The ELK services are deployed using Docker containers to simplify installation, improve portability, and provide a consistent execution environment.

Finally, Kibana provides an interactive visualization layer for monitoring security events. It connects directly to Elasticsearch and displays dashboards that summarize attack statistics, event timelines, log trends, and system activities. These dashboards enable security analysts to identify suspicious behavior, investigate incidents, and monitor network security in real time. By integrating centralized log management with interactive visualization, the ELK Stack forms the foundation of the proposed cloud-based security analytics framework.

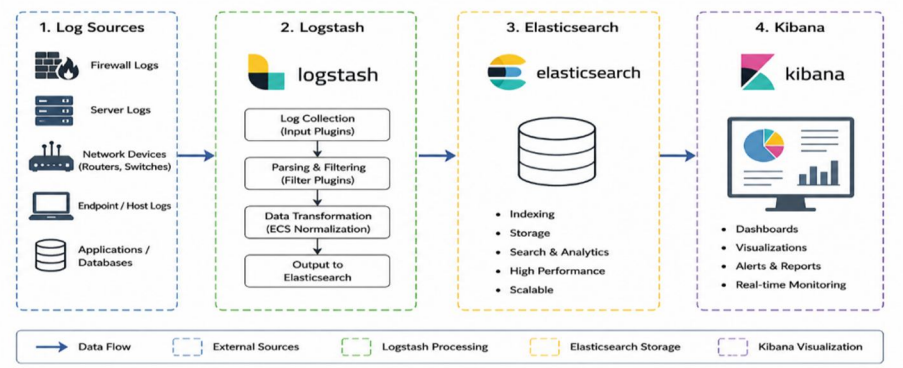


Figure 3. ELK Stack-Based Log Processing Workflow

D. Machine Learning Model Development

The machine learning module is designed to classify network traffic into benign and malicious categories using supervised learning techniques. After preprocessing, the CICIDS2017 dataset containing 2,497,980 records and 70 numerical features was used for model development. The dataset was divided into 80% training and 20% testing subsets using stratified sampling to preserve the original class distribution. Before training, feature values were standardized using the StandardScaler algorithm to ensure consistent feature scaling and improve model convergence. Three ensemble-based machine learning algorithms were implemented and evaluated: Random Forest (RF), XGBoost, and LightGBM. Random Forest constructs multiple decision trees using bootstrap aggregation to improve classification robustness. XGBoost employs gradient boosting with optimized tree learning, making it effective for handling high-dimensional network traffic data. LightGBM utilizes a leaf-wise tree growth strategy with Gradient-based One-Side Sampling (GOSS) to achieve faster training while maintaining high predictive performance. The models were trained sequentially to reduce memory consumption and support efficient execution on systems with limited hardware resources.

To address the significant class imbalance present in the CICIDS2017 dataset, each model applied an imbalance handling strategy during training. Random Forest used class_weight='balanced_subsample', XGBoost applied balanced sample weights, and LightGBM used its built-in is_unbalance option with GOSS sampling. After training, the models were evaluated using Accuracy, Precision, Recall, and Macro F1-score. The best-performing model was subsequently selected for integration into the proposed SIEM framework for automated attack prediction and real-time security analytics.

IV. RESULTS AND DISCUSSION

A. Traditional Security Information and Event Management Systems

The proposed SIEM framework was implemented using Python 3.11 and the ELK Stack, comprising Elasticsearch 8.11, Logstash 8.11, and Kibana 8.11. The machine learning models were developed using Scikit-learn, XGBoost, and LightGBM libraries. Docker Desktop was used to deploy the ELK services, providing a consistent and scalable execution environment. The experiments were conducted using the CICIDS2017 dataset, which was divided into 80% training and 20% testing sets. Model performance was evaluated using Accuracy, Precision, Recall, and Macro F1-score.

Parameter	Value
Programming Language	Python 3.11
Dataset	CICIDS2017
Records	2,497,980
Features	70
Attack Classes	15
Train-Test Split	80:20
ML Models	Random Forest, XGBoost, LightGBM
SIEM Platform	ELK Stack
Deployment	Docker Desktop
Evaluation Metrics	Accuracy, Precision, Recall, Macro F1-score

Table I. Experimental Configuration

B. Performance Comparison

Three supervised machine learning models were evaluated using the same preprocessed dataset and experimental configuration. Random Forest, XGBoost, and LightGBM were trained independently and assessed using standard performance metrics. The experimental results demonstrate that all three models achieved high classification accuracy. XGBoost achieved the highest Macro F1-score, while LightGBM achieved the highest overall accuracy. Considering the balance between accuracy and multiclass classification performance, XGBoost was selected as the prediction model for integration into the proposed SIEM framework.

Model	Accuracy	Macro F1 Score
Random Forest	99.53%	84.41%
XGBoost	99.53%	89.57%
LightGBM	99.87%	89.31%

Table II. Performance Comparison of Machine Learning Models

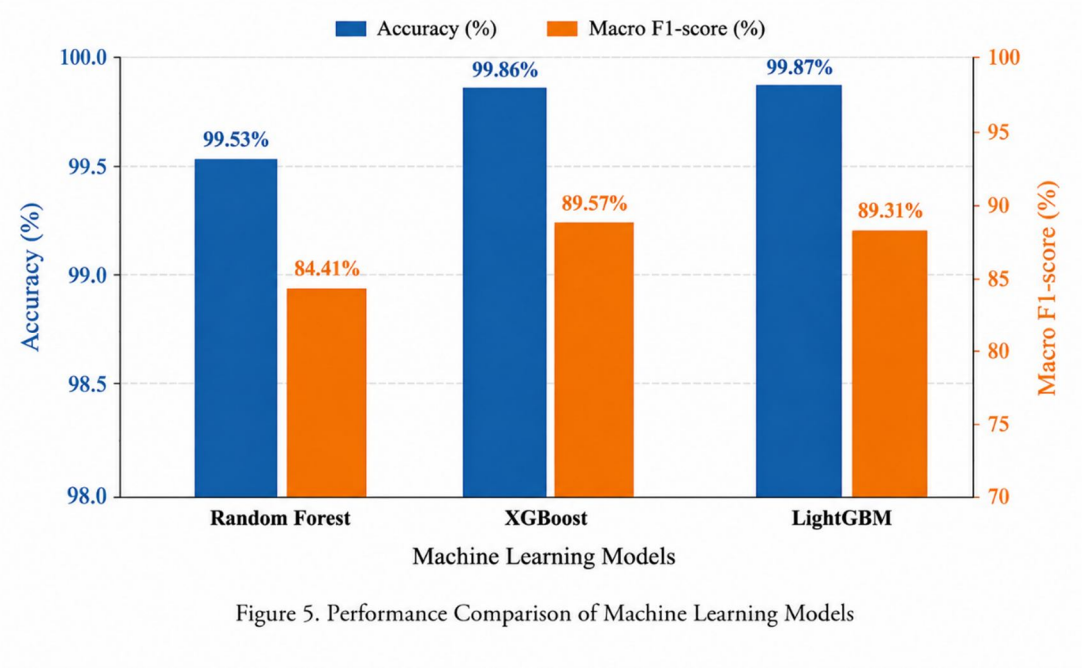


Figure 5. Performance Comparison of Machine Learning Models

C. Discussion

The experimental results demonstrate that ensemble learning techniques are highly effective for multiclass intrusion detection using the CICIDS2017 dataset. Random Forest provided reliable classification performance with reduced computational complexity, whereas XGBoost and LightGBM achieved superior predictive performance through gradient boosting techniques. Although LightGBM achieved slightly higher overall accuracy, XGBoost produced the highest Macro F1-score, indicating better classification performance across minority attack classes. Based on this balanced performance, XGBoost was selected as the final prediction model for integration into the proposed cloud-based SIEM framework. The integration of ELK Stack with machine learning enables centralized log management, automated attack prediction, and real-time visualization, thereby enhancing the efficiency of security monitoring and incident response.

V. CONCLUSION

This paper presented a Cloud-Based Security Analytics SIEM Using ELK Stack and Machine Learning for intelligent cyberattack detection and centralized security monitoring. The proposed framework integrates the ELK Stack with supervised machine learning techniques to automate log collection, processing, attack detection, and visualization. The CICIDS2017 dataset was used to train and evaluate Random Forest, XGBoost, and LightGBM models. Experimental results showed that XGBoost achieved the best overall performance with 99.86% accuracy and a Macro F1-score of 89.57%, making it the most suitable model for deployment in the proposed SIEM framework.

The developed system successfully combines cloud-based log analytics with machine learning to provide centralized monitoring, automated threat detection, and real-time visualization through Kibana dashboards. The proposed framework offers a scalable and efficient solution for enhancing cybersecurity monitoring and supporting Security Operations Centers (SOCs). Future enhancements may include integrating live network traffic analysis, deep learning-based detection models, and automated incident response to further improve detection capabilities.

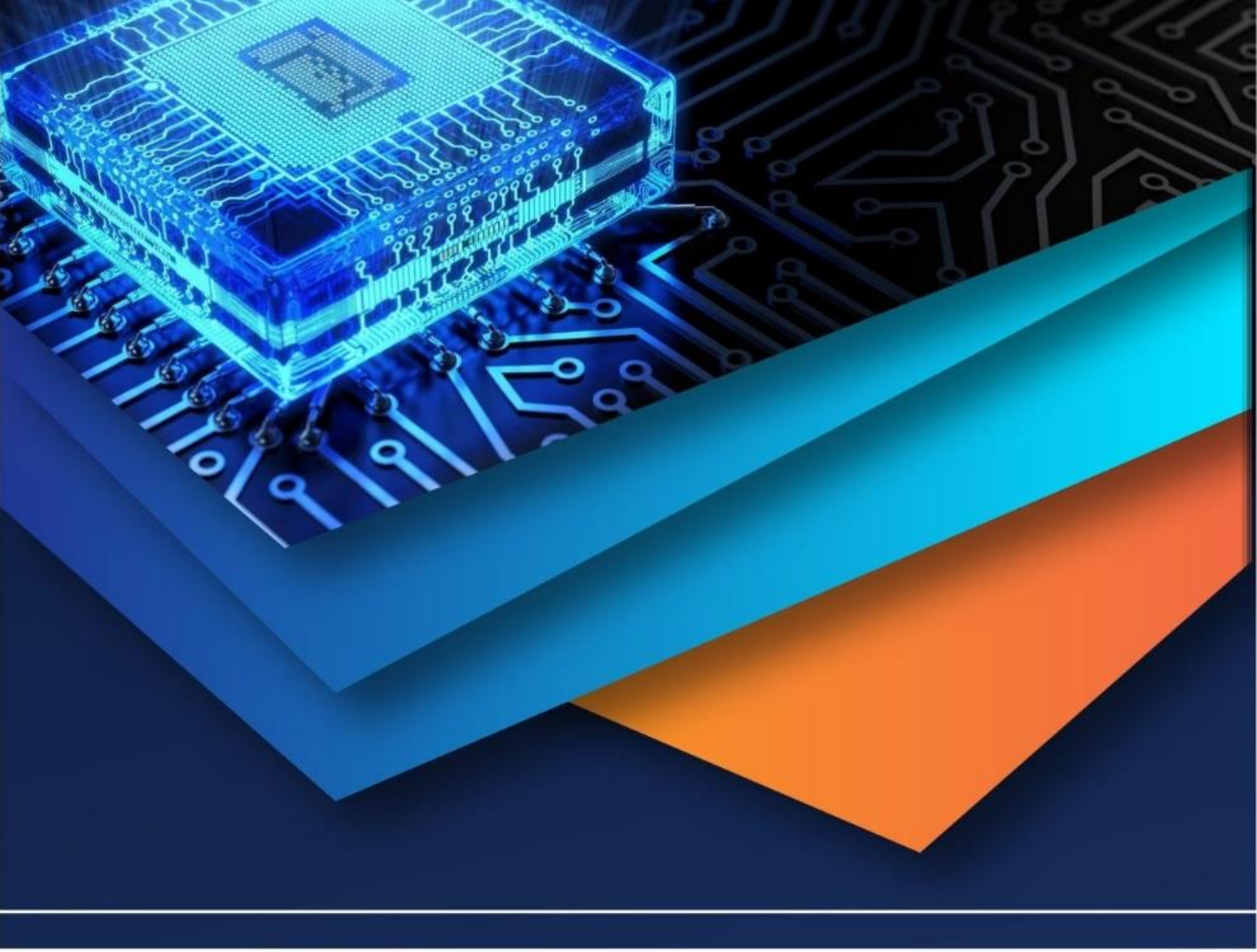
VI. ACKNOWLEDGMENT

I express my sincere gratitude to my guide, Dr. P. Sammulal, Professor, Jawaharlal Nehru Technological University Hyderabad, for his valuable guidance, continuous support, and constructive suggestions throughout the development of this research work. His encouragement and technical insights greatly contributed to the successful completion of this project.

I also extend my heartfelt thanks to the faculty members of the Department of Computer Science and Engineering for their support and motivation during the course of this study.

REFERENCES

[1] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP), Madeira, Portugal, pp. 108–116, 2018.
[2] A. H. Lashkari, G. D. Gil, M. S. I. Mamun, and A. A. Ghorbani, “Characterization of Tor Traffic Using Time Based Features,” Proceedings of the 3rd International Conference on Information Systems Security and Privacy (ICISSP), Porto, Portugal, pp. 253–262, 2017.
[3] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp. 785–794, 2016.
[4] G. Ke, Q. Meng, T. Finley et al., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 3146–3154, 2017.
[5] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[6] B. Burns, B. Grant, D. Oppenheimer, E. Brewer, and J. Wilkes, “Borg, Omega, and Kubernetes,” Communications of the ACM, vol. 59, no. 5, pp. 50–57, May 2016.
[7] Elastic NV, Elasticsearch: The Official Distributed Search and Analytics Engine, Available: https://www.elastic.co/elasticsearch
[8] Elastic NV, Logstash: Data Collection and Processing Pipeline, Available: https://www.elastic.co/logstash
[9] Elastic NV, Kibana: Data Visualization and Analytics Platform, Available: https://www.elastic.co/kibana
[10] F. Chollet, Deep Learning with Python, 2nd ed., Manning Publications, 2021.
[11] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed., Pearson Education, 2021.
[12] W. Stallings, Network Security Essentials: Applications and Standards, 7th ed., Pearson, 2022.
[13] D. E. Denning, “An Intrusion Detection Model,” IEEE Transactions on Software Engineering, vol. SE-13, no. 2, pp. 222–232, Feb. 1987.
[14] NIST, Guide to Intrusion Detection and Prevention Systems (IDPS), NIST Special Publication 800-94, National Institute of Standards and Technology, 2007.
[15] C. Krügel, F. Valeur, and G. Vigna, Intrusion Detection and Correlation: Challenges and Solutions, Springer, 2005.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY