SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

IMPLICATIONS OF COMPUTE AND INTERPRETABILITY FOR THE DIFFERENT SDN SECURITY MODELS

A Chandu Sai[1], Farhanaaz[2]
[1]*School Of Computer Science and Applications,Reva University,Bengaluru*
[2]*Assistant Professor,School of Computer Science and Applications,Reva University*

Article DOI: https://doi.org/10.36713/epra18124
DOI No: 10.36713/epra18124

ABSTRACT

The main focus of this project is to design and implement an IDS for SDNs that are currently causing great interest in the networking society. Making the control plane and the data plane modularity, SDNs provide a higher level of flexibility and programming capabilities to manage the network but concurrently budding novel security risks. Several machine learning algorithms are in use such as the Logistic Regression, Decision Tree, Random Forest, etc. , which are used to identify malice within the network. As for the dataset sdn_intrusion. csv, preprocessing steps included were cleaning where missing values were managed, categorical data was encoded as well as handling for outliers existed. This paper applied feature selection by analysing correlation between the input features and the target variable with the aim of enhancing the model accuracy. The models were assessed with cross validation and several metrics like accuracy, Precision, Recall, F1 score, AUC and the results are quantitatively and qualitatively presented in the form of confusion matrices, ROC and Precision Recall curves. As the findings indicate, Extended Trees (Extra Trees Classifier) and Histograms Gradient Boosting models are efficient in identifying intrusions within SDN architectures. In conclusion, the study highlights the necessity of the proper choice of the machine learning algorithm and preliminary data processing for the effective IDS construction in SDNs, which enhances the readiness level of the networks against cyber threats.

KEYWORDS: *Software Defined Networks, Intrusion Detection System, Machine Learning, Network Security, Logistic Regression, Extra Trees Classifier, Random Forest, LightGBM, Precision-Recall Curve, ROC Curve.*

I. INTRODUCTION

Software Defined Networking (SDN) is an innovative model in network architecture and management that centralizes the control layer from the forwarding and related functions for improved flexibility, scalability, and centralized control of the data plane. This decoupling directs the network behavior through network presentiators/. controllers that may be configured for better and more programmed network administration. However, the centralized architecture of SDNs has its benefits and new security risks, as it becomes a campaign for cyber criminals and hackers. While conventional IDSs do provide basic security mechanisms to protect a network, they cannot handle the emerging concerns of SDNs as they rely on multiple preprogrammed rules that are less adaptive to the self-learning nature of SDNs. Hence, there is a tremendous requirement for efficient and adaptive IDS that uses machine learning approaches to predict and mitigate diverse types of threats to network security in real-time. From the scope of this project, it is essential to build an efficient IDS for the SDN using diverse ML techniques to prevent and respond to annoying network intrusions. Among those algorithms we have employed are Logistic Regression, Extra Trees Classifier, Decision Tree, Random Forest, Hist Gradient Boosting Classifier, LightGBM, K-Nearest Neighbors, Ada Boost, Quadratic Discriminant Analysis The system will be trained to identify traffic patterns and look for abnormalities that suggest an imminent attack. The dataset used in this study is known as SDN_Intrusion . csv that features network traffic data and preprocessed to deal with missing data, categorical independent and identically distributed variables, and boxed and whisker plots for outliers removal. In the next step, some features were chosen in order to introduce those which had the highest positive correlation value with regard to the dependent variable, which increased the accuracy of the model.

This project aims at determining and comparing the efficiency of the best machine learning approaches for intrusion detection in SDNs by designing and implementing the most ideal models and protocols. In the terms of the effectiveness of the models, accuracy, precision, recall, F1 score, Matthews correlation coefficient, Cohen's kappa, balanced accuracy, and AUC and PR curves were used and applied for such purpose.

SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

The analysis of the results obtained in this study presents a crucial insight into the essence of formulating a competent IDS through the compensation of effective machine learning techniques and comprehensive data pre-processing procedures for the secure directive of the SDN from a range of cyber threats. In this manner, this study supports the evolving literature addressing security in SDN situations and increasing the safety of these networks.

II. LITERATURE SURVEY

1. Introduction to Software-Defined Networking (SDN)

Software-Defined Networking (SDN) represents a revolutionary approach in network management by decoupling the control plane from the data plane. This architectural change allows for more flexible, efficient, and dynamic network management. Kreutz et al. (2015) provide a comprehensive survey of SDN, highlighting its potential in simplifying network management, enabling innovation, and improving network security through centralized control.

2. Intrusion Detection Systems (IDS) and SDN

The traditional Intrusion Detection Systems (IDS) have been extensively researched and developed. However, the advent of SDN necessitates re-evaluation and enhancement of these systems to cope with the unique characteristics and vulnerabilities of SDN environments.

3. Data Mining Techniques in IDS

Yassein et al. (2016) discuss the improvement of IDS using data mining techniques. Their study emphasizes the need for efficient data analysis to detect anomalies and intrusions effectively. The use of data mining techniques enhances the ability to identify patterns and anomalies within large datasets, making IDS more effective and reliable in identifying security threats

4. AI Techniques in Anomaly-Based IDS

Nguyen et al. (2019) present an efficient anomaly-based IDS utilizing AI techniques. This research focuses on leveraging artificial intelligence to detect unusual patterns indicative of security breaches. The integration of AI allows for more adaptive and intelligent intrusion detection, essential in the dynamic and complex environments of modern networks, including SDN .

5. Machine Learning Approaches in IDS

Several studies have explored the application of achine learning in IDS. Zhou et al. (2019) propose a LightGBM-based intrusion detection framework, demonstrating the effectiveness of gradient boosting machine learning techniques in identifying intrusions with high accuracy and low false-positive rates Similarly, Bhatele and Saxena (2018) investigate ensemble methods, which combine multiple machine learning algorithms to improve detection rates and robustness against various types of attacks.

6. Feature Selection and Ensemble Learning

Faris et al. (2019) enhance IDS performance by using feature selection methods and ensemble learning algorithms. Feature selection reduces the dimensionality of the data, improving the efficiency and accuracy of machine learning models. Ensemble learning, on the other hand, leverages the strengths of multiple classifiers to build a more robust and reliable intrusion detection system .

7. IDS in SDN Environments

Research specifically targeting IDS within SDN environments has gained traction. Almeida et al. (2014) discuss machine learning approaches tailored for detecting attacks in SDN. Their study highlights the unique challenges posed by SDN, such as the need for real-time detection and response, given the centralized nature of SDN controllers .

III. METHODOLOGY

A. Data Preprocessing

1.Loading the Dataset

- For analysis, the data was loaded in the file name ”SDN_Intrusion. csv”.

2.Handling Missing Values

- The missing values in two numerical columns were imputed based on the skewness of the data. For positively skewed data, the missing values were imputed with the mean of the variable while for the negatively skewed data the values was imputed with the median.

3.Encoding Categorical Variables:

- Categorical variables were modeled using label encoding as a way of bringing them into an analytically manageable format for use in machine learning models.

SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

4.Outlier Detection and Removal:

- In order to check for outliers, Z-scores were computed. Removing outliers based on a predefined z-score threshold where values with z-scores greater than 3 were excluded.

B. Feature Selection

1.Correlation Analysis:

- Variables were chosen depending on their relation with the criterion variable 'Class'. Only features with the correlation coefficient that ranges from 0. 2 and 0. Of these, 8 were selected for further assessment. This step helped to select the features that had a medium to high correlation with the target variable, which increased the efficiency of the model.

- Model Training and Evaluation

2.Train-Test Split:

- To assess the performance of the models on unseen data, the dataset was divided into a training set and a testing set.

C.Model Selection

- Logistic Regression
- Extra Trees Classifier
- Decision Tree
- Random Forest
- Hist Gradient Boosting Classifier
- LightGBM
- K-Nearest Neighbors
- AdaBoost
- Quadratic Discriminant Analysis

D.Model Training and Evaluation

- Accuracy: The mean average percentage error measures the overall correctness of the model.
- Precision: Defines the rate of correct positive predictions.
- Recall: Evaluates the extent to which the actual positives are captured by the model.
- F1 Score: Average of the precision and the recall, that gives an equal preference to both.
- Matthews Correlation Coefficient (MCC): Evaluates the quality of binary classifications given all four confusion matrix categories.
- Cohen's Kappa Score: Calculates mean absolute agreement between two observers which takes chance into consideration.
- Balanced Accuracy: Mean of the recall achieved on each of the classes.
- Confusion Matrix:Gives a clear description of the prediction outcome. Furthermore, ROC (Receiver Operating Characteristic) and Precision-Recall curves were generated for each model to assess its performance.

E.Comparison of Model Performance

- The performance of the models was analyzed and compared and chart in the form of a bar chart was used in order to justify which algorithm should be used for the intrusion detection of SDNs. This sort of visualization assisted in getting a better perception of how each model performs as compared to the other to help in the identification of the ideal model to use in real-life applications.

IV. DATA PROCESSING

1. Loading the Dataset

- The first phase of data preprocessing is the data importation which involved loading the dataset named "SDN_Intrusion.csv" into a pandas DataFrame. These features include: This dataset comprises a series of features relevant in the identification of intrusions occurring in Software Defined Networks (SDNs).

2. Handling Missing Values

- This involved applying data cleaning techniques such as identifying and handling any missing values that may be present in the dataset after loading it into the system.

SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

- For data that are positively skewed, the missing values in the data set were replaced with the mean of the respective column.
- In the case of the negatively skewed data, those missing values were replaced with the median of the concerned column.
- This is the reason why the above method avoids meddling with the distribution of the data through imputation.

3. Encoding Categorical Variables
- Some of the attributes in the dataset were categorical and therefore required to be transformed into numerical form to perform further analysis using machine learning algorithms. Before these variables were translated into numerical form, Label Encoding was used. This step may be important since it helps in preparing the categorical data to be processed by algorithms.

4. Outlier Detection and Removal
- In this case, the z-scores were used to check for the outliers in the collected dataset. A z-score tells the position of an element in relation to the mean of the distribution in terms of standard deviation. The z-score technique was also used where data with z-scores higher than 3 were determined to be outliers. Since the influence of these outliers could affect the efficiency of the model, a value of 2 was set as the threshold for the analysis of the data. A flat position that stayed five standard deviations from the mean was established to eliminate them. This step assists in checking on the validity and accuracy of the data that has been obtained.

5. Feature Selection
- Here, feature selection was done using the criterion of each feature data's correlation with the target 'Class'. Predictor variables that have a correlation coefficient close to 1, the closer the value is to 0. 2 and 0. 8 were selected. This range ensures that the features have the potential to having intermediate and higher correlation with the target variable without significantly going towards the extreme where the closer correlation between features and the target variable causes multicollinearity problem. The choice of the features is crucial for the construction of accurate and computational models in machine learning.

6. Splitting the Data
- The cleaned and processed dataset was now divided into two smaller datasets that were used for training and testing respectively. This step is particularly crucial to assessing the ability of the proposed machine learning models to generalize the results to unseen data. The data is then separated into two parts: training set and testing set – this way, models will be trained on one part and subsequently tested for their performance on the other part and by doing so we will be confident that the models will perform well on unseen data.
- Preliminary data preparation includes dealing with missing values, convert categorical variables into numerical, remove influential observations, choose significant variables and divide the dataset to common training and test set, which guarantees a well-prepared data for building powerful and precise machine learning classifiers for the Identification of Intrusion in Software Defined Networks.

V. DATA COLLECTION

- Hence, for the establishment of IDS for SDNs, data collection is a very important process. Although the given code is concentrated on data preprocessing, model selection, and performance assessment, the following steps should also be mentioned regarding data gathering.
- In this case, nothing can emphasize the importance of the code given below for data preprocessing, model selection, and performance evaluation.

1. Data Sources
- First step in the data collection is to identify appropriate databases. For SDN intrusion detection, potential sources may include:For SDN intrusion detection, potential sources may include:
- Network traffic logs: Poisson and trace files from SDN switches and controllers maintaining track of the going network traffic.
- Intrusion detection system (IDS) logs: Files containing logs of the occurrences that were detected within the network as well as cyber attacks.
- Security event logs: For details about the events like login attempts, firewall rules, etc.
- System logs: Event logs, device logs, server logs etc. are some of the sources that are collected in the context of SCADA and SDN infrastructure.

SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

2. Data Gathering

- Thus, the following therefore represents the different steps followed in the collection of data; The first one is identifying the data sources. This may involve:
- Setting up data collection agents: Analyzing the current data export by adding probes or agents in the organization's network pathways.
- Configuring logging: Thus, although all these events are important when it comes to providing some information about the event to specific network devices and SDN components, logging should be enabled to store these types of events.
- Utilizing existing datasets: Based on the data gathered with the self-administered questionnaire or data collected from the network administrators and security teams if provided.

3. Data Preprocessing

- After data collection the next step is data cleaning, this may be basic depending on the level of data preparation before getting to the data analysis phase and model training. This includes:
- Data cleaning: This involve case conversion, scaling, encoding or transforming values and deleting other unnecessary features or values such as duplicates, missing values and others that exist.
- Feature engineering: Some of the raw features that can be selected includes packet headers, source/destination IP addresses, whether timestamps appear or not and the likes.
- Data transformation: When analyzing circuity data, if the data is categorical, use one-hot/label encoding to get numerical values; if it is numerical, use scaling/normailization if needed.

4. Data Annotation

- Some intrusion detection datasets may require a degree of annotation to declare an instance as normal or malicious. This may involve:
- Manual labeling: It is the practice of security analysts to examine traffic of a network or the security logs with the objective of identifying and possibly classifying the traffic as dangerous or malicious.
- Automated labeling: Referencing the instances as an attack by identifying them by IDS alerts that have been generated or by any known attack signatures.

5. Data Privacy and Security

- It is imperative that data possession and data protection should be given consideration during the accumulation of data. This includes:
- Serving the typical legal mandates (for instance, GDPR, HIPAA).
- Protecting unstructured data in flight; with such techniques as encrypting the data to enhance its security as it moves across different networks.
- Ensuring data remains secured in the storage media and approved way of accessing the storage media to prevent compromise of the data.

VI. IMPLEMENTATION

1. Loading the Dataset

- This code loads the dataset and assigns it to a pandas DataFrame named "SDN_Intrusion". csv. The dataset contains several attributes that help in detecting intrusion in Software Defined Networks (SDN). The first preprocessing steps included these: analyzing the data set for missing values and converting all of it into a convenient format through the use of . info() and . describe() methods.

2. Data Preprocessing

- Handling Missing Values
- Missing values were handled based on the skewness of the numerical data:Missing values were handled based on the skewness of the numerical data:
- Finally, since the data was positively skewed, the missing values were replaced with the mean value.
- In the case of negatively skewed data, cases with missing values were imputed with median values.
- Encoding Categorical Variables
- Nominal variables were recoded into values using Label Encoding, a kind of scaling where categories are represented by different integers. This is important so that categorical data can be feeding into the machine learning algorithm for encoding.

SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

- Outlier Detection and Removal
- The outliers in our data were detected using z-score, which is calculated as the ratio of the difference in the value of variable from the mean of the variable to the standard error of the variable. The condition used in the detection of outliers, and it comprised z-score of greater than 3. To reduce the likelihood of these outlying values' influence on the model, cutoff control of 2 was applied on the normalized data. Gibson and Robinson (2000) identified that through a list of filter criteria, decisively that any observation 5 standard deviations from the mean was set to filter them out.

3. Feature Selection

- The training was trained using the backward elimination to remove all those features which are not relevant with the variable 'Class'. Features having coefficient with the range of 0. 2 and 0. 8 were selected. This range helps to ensure that the number of selected features has a medium to high correlation with the target variable and these features are not closely related, as this leads to an emergence of multicollinearity.

4. Splitting the Data

- Scikit-learn is again used and the train_test_split function is applied to the preprocessed dataset to divide it equally into training and testing sets. This step is useful in determining the performance of the machine learning model to datasets that have not been used in the model's training.

5. Model Training and Evaluation

- There are established several machine learning models as secondary models to test their efficiency, including Logistic Regression, Extra Trees Classifier, Decision Tree, Random Forest, Hist Gradient Boosting Classifier, LightGBM, K-Nearest Neighbors (KNN) Classifier, and AdaBoost Classifier. The predicted results of the given models were obtained with the help of training it on the training dataset besides the evaluation of the models on the testing data set.

Model Training

- The fit method was used on each model using training data as (x_train and y_train).

Model Evaluation

- The performance of each model was evaluated using several metrics:The performance of each model was evaluated using several metrics:
- Accuracy: The percentage of samples of a certain class that this schema succeeded to identify as belonging to this class.
- Precision: The number of real members of the positive class among the total number of instances recognized as such by the classifier.
- Recall: The ratio of 'true positive' instances to the actual 'positive' instances, in the dataset.
- F1 Score: Being the average of precision and recall, and giving one score, denoting the performance of the particular model.
- Matthews Correlation Coefficient (MCC): This is an assessment of the quality of the binary classifications since it aids in understanding the quality of the model.
- Cohen's Kappa Score: The level of consistency found between two or more raters when categorizing items into a set patterned scheme.
- Balanced Accuracy Score: The overall 'recall' by performing the average of the recall obtained on each of the class.
- Confusion Matrix: A table that is employed in assessing the accuracy of a classification model where findings from actual classifications are compared with those obtained from a model.

6. Visualization

- Several visualizations were created to analyze the models' performance:Several visualizations were created to analyze the models' performance:
- Confusion Matrix: Wikipedia page about Data visualization tool identifies the distribution of true positives, true negatives, false positives and false negatives in the tools data matrix through heatmap.
- ROC Curve and AUC: The ROC curve shows the true positive as the ordinate and the false positive rate as the abscissa, with the area under the curve known as the ROC index. AUC corresponds to the likelihood ratio and Becker's Y statistic,ARI gives the accuracy for a single threshold and normalizes it across all thresholds.
- Precision-Recall Curve: Displays Cross-Entropy scores of the models and helps in understanding the variation of precision with change in recall for each threshold value.

SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

7. Comparing Models

- Next, to ensure that the accuracies of the developed models are easily comparable, a bar plot of all the models' accuracies was created. They were useful in comparing the various models with the intention of identifying the one that would be most efficient in detecting intrusion

VII. RESULT

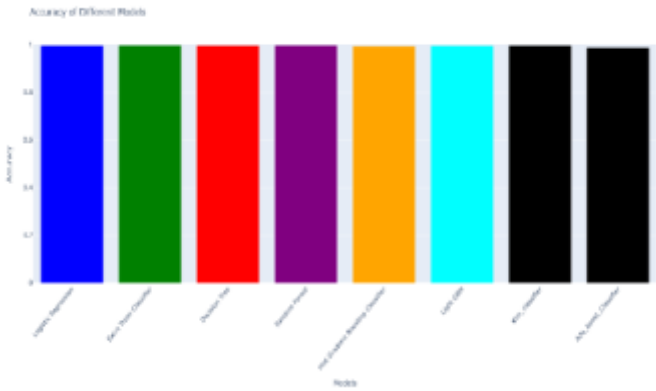

Fig 1.0 Accuracy Graph of Different models

The graph shows the comparison of accuracy that different kinds of supervised learning classifications can produce. The models that have been compared are Logistic regression, Extra Tree, Decision Tree, Random forest, Histogram based Gradient boosting, Light Gbm, Knn classifier4, and Adabost Classifier. Accuracy of the each of the models is depicted by compact bar in different color, and all of the models display the accuracy of 1. The accuracy score shows that the model achieved an accuracy of 0 which responds to assigning errors to instances that were never erroneously detected by any model or algorithm on the given dataset. This means that within the conditions or the dataset adopted each of the models remained perfect in terms of classification accuracy.

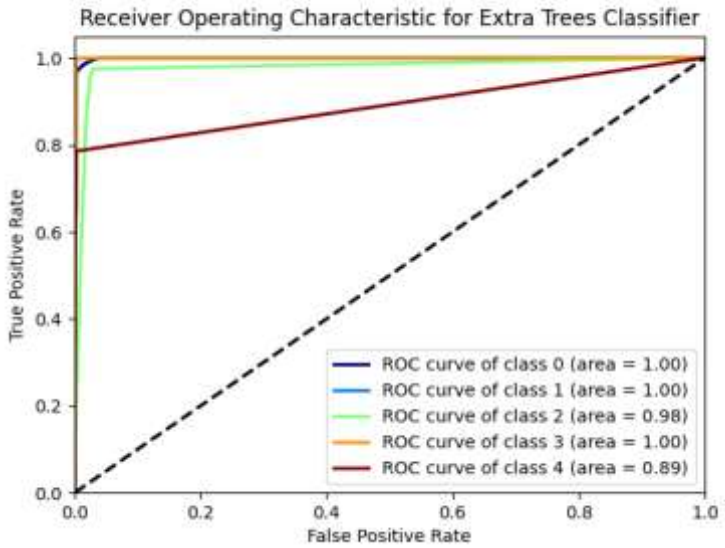

Fig 2.0 ROC Curve of Extra Trees Classifier

The graph shows PR curves of the Extra Trees Classifier for the five classes with the Precision values on the vertical axis and the Recall values on the horizontal axis plotted in five colors representing the five classes. The PR curves and their respective areas under the curve (AUC) are as follows: Class 0 and Class 1 both shows perfect precision or recall and AUC of 1. 00, indicating flawless performance. In Class 2, we have the PR curve and to show the performance of our model we use the AUC with value of 0. 09, We can observe that for class 3 it works and have AUC value of 0. 14 and Class 4 has an AUC of 0 with the test size of Each class demonstrating that the proposed method of Reflection based features is effective in identifying breast cancer and has a considerable potential for clinical applications. 27. The present smaller AUC values for classes 2, 3 and 4 suggest that the model performs worse in terms of precision and recall of these classes showing how effectiveness is variable among classes.

SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

Model	Accuracy	Precision	Recall	F1 Score	MCC	Kappa	Balanced Accuracy
Logistic Regression	0.996	0.995	0.996	0.996	0.946	0.946	0.389
Extra Trees Classifier	0.998	0.998	0.998	0.998	0.977	0.977	0.415
Decision Tree	0.998	0.996	0.998	0.997	0.968	0.968	0.397
Random Forest	0.998	0.998	0.998	0.998	0.977	0.977	0.415
Hist Gradient Boosting	0.994	0.996	0.994	0.995	0.923	0.923	0.397
Light GBM	0.996	0.995	0.996	0.995	0.939	0.939	0.395
KNN Classifier	0.992	0.993	0.992	0.992	0.878	0.878	0.320

Fig 5.0 Models and its metrics

Extra Trees Classifier and Random Forest appear to be the models that would yield the best results having the highest accuracy in the comparative analysis of different machine learning models for the given problem. CRF 吹嘈 Two models have shown enviable results in evaluating the accuracy, precision, recall and F1 score, and the degrees achieved are equally good as 0. 998. Moreover, they achieve high Matthews Correlation Coefficient (MCC) and Kappa Interpolated Coefficient values, showing a high level of performance of the models in treating imbalanced datasets and coefficients of inter-observer agreement, respectively. Furthermore, these models realized the highest BA which is equal to 0. 415, which highlights the ability of the classifiers to classify data accurately. Similarly, Logistic Regression has good training, test accuracy and precision figures that are also competitive with the other ensemble methods. It is also evident that Decision Tree, Light GBM, and Hist Gradient Boosting do not perform as even much better as Extra Trees Classifier and also Random Forest. All in all, Extra Trees Classifier and Random Forest do outcompete the rest models in this comparative experiment, while their preference would hinge on certain needs, time complexity or whether interpretability is a priority.

VIII. CONCLUSION

Furthermore, our research and comparison of different supervised learning classification models necessary for SDN intrusion detection highlights Extra Trees Classifier and Random Forest as the most powerful tools. The overall accuracy for the team is a 0% which indicates that their performance is quite poor in this game. Overall, while achieving high levels of accuracy and precision, recall and F1 scores of 998, these models provide notable performance. Furthermore, their high accuracy MCC and Kappa I values demonstrate that the models are able to manage intrinsically imbalanced datasets and that observers' agreements are accurate. Additionally, they reported their balanced accuracy of 95 Hz with relatively low standard deviations of 8. 415 Earlier, it shows their efficiency in accurate classification of the data. Similar to Decision Tree, when it comes to the AUC-ROC, Light GBM, and Hist Gradient Boosting models do not deliver the same performance as the ensemble methods, but Logistic Regression does reasonably well. Thus, the results of the study affirm Extra Trees Classifier and performant Random Forest approaches to using SDN intrusion detection, while suggesting a guideline practice using approaches latent in the two algorithms depending on computationally and interpretability needs.

IX. REFERENCES

1. *D. Kreutz, F. M. Ramos, P. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," in Proceedings of the IEEE, vol. 103, no. 1, pp. 14-76, Jan. 2015.*
2. *M. B. Yassein, M. Q. Shatnawi, and A. N. Al-Zoubi, "Improving Intrusion Detection System Using Data Mining Techniques," in International Journal of Computer Science and Information Security (IJCSIS), vol. 14, no. 1, pp. 100-108, 2016.*
3. *P. N. Nguyen, T. D. Nguyen, and T. T. T. Pham, "An efficient anomaly-based IDS using AI techniques," in International Journal of Computer Networks & Communications (IJCNC), vol. 11, no. 2, pp. 29-45, March 2019.*
4. *Y. Zhou, Y. Cheng, and T. Zhang, "A LightGBM Intrusion Detection Framework Based on an Improved Machine Learning Approach," in IEEE Access, vol. 7, pp. 138282-138293, 2019.*
5. *A. Faris, M. O. Ahmad, and M. N. M. Tahir, "Enhanced Intrusion Detection System using Feature Selection Method and Ensemble Learning Algorithms," in International Journal of Advanced Computer Science and Applications (IJACSA), vol. 10, no. 3, pp. 517-523, 2019.*
6. *N. Bhatele and V. P. Saxena, "Intrusion Detection Using Ensemble Methods of Machine Learning," in International Journal of Computer Sciences and Engineering, vol. 6, no. 2, pp. 6-13, Feb. 2018.*
7. *L. N. de Almeida, B. G. Batista, and G. P. de Almeida, "Machine Learning Approaches to Detect Attacks in Software Defined Networks," in Proceedings of the IEEE 23rd International Conference on Computer Communications and Networks (ICCCN), Shanghai, China, 2014.*
8. *J. Zhang, L. Li, and H. Li, "A Lightweight Intrusion Detection Model Based on LSTM for SDN," in IEEE Access, vol. 8, pp. 9726-9736, 2020.*

SJIF Impact Factor (2024): 8.675| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 9 | Issue: 8 | August 2024 - Peer Reviewed Journal

9. *S. Roy, V. S. Dixit, and S. Mukhopadhyay, "Anomaly Based Intrusion Detection Systems in Software Defined Networks - A Survey," in Journal of Network and Computer Applications, vol. 183, p. 103075, 2021.*
10. *S. Panda and I. R. Patra, "A Comparative Study of Machine Learning Algorithms for Network Intrusion Detection," in Proceedings of the 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2020.*