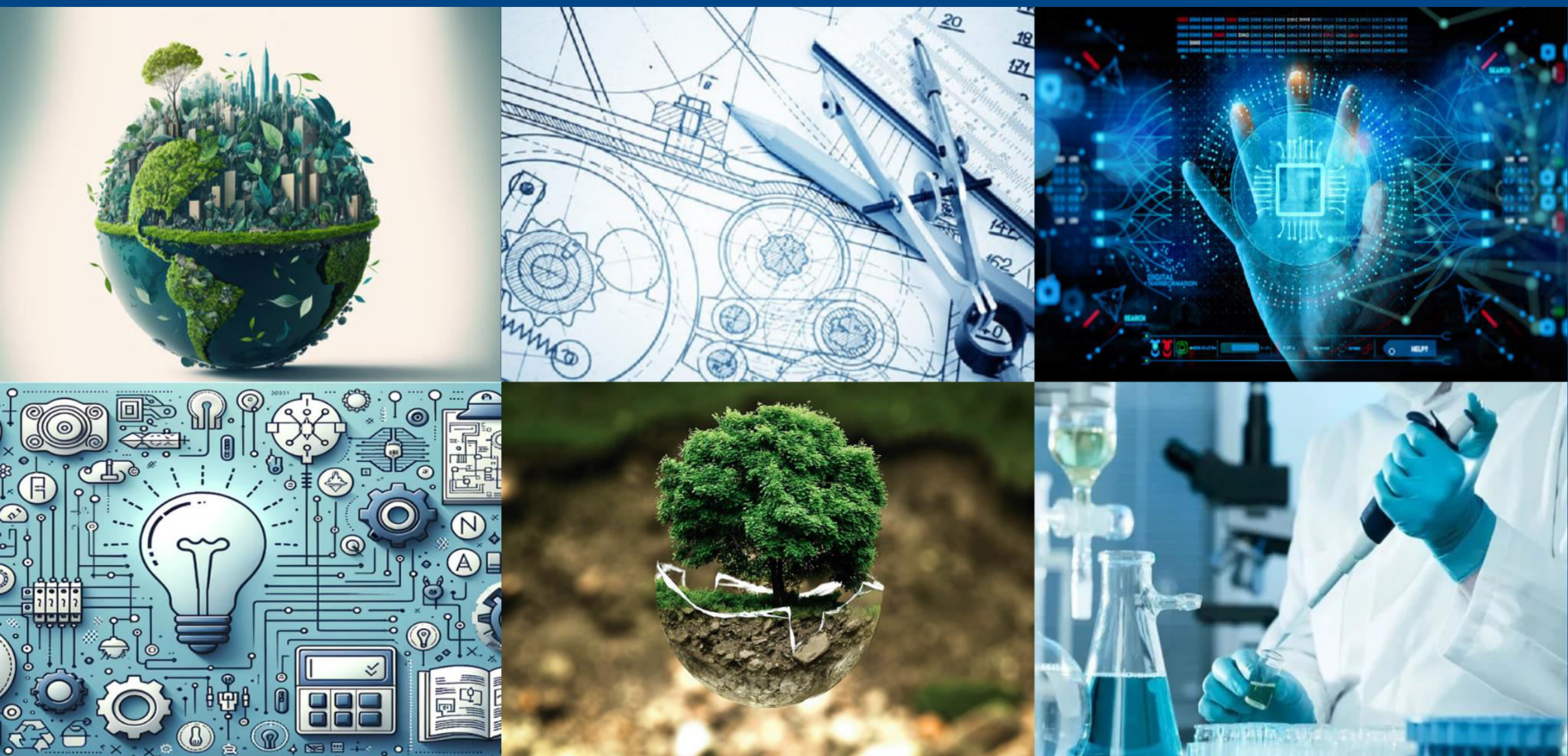
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 4, April 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Predicting Cyberattack using Machine Learning

Dr.B.Leelavathi, Mr. N.Manoj kumar

Associate Professor, Department of Computer Technology, Dr.N.G.P. Arts and Science College, Coimbatore, India

Student, Department of Computer Technology, Dr.N.G.P. Arts and Science College, Coimbatore, India

ABSTRACT: As the incidence of cyber data breaches continues to increase, conventional manual investigation techniques for tracing cyber-attacks become more time-consuming and prone to errors. Advances in cyber threats, which tend to follow similar patterns repeatedly, pose difficulty in investigating them in time. Cyber-attacks, carried out through cyberspace, tend to destabilize, cripple, manipulate, or compromise an organization's computing resources, threatening data integrity as well as facilitating unauthorized access to confidential data. The dynamic nature of cyberspace brings with it concerns regarding the future of the internet, especially with its growing number of users. New paradigms, including big data created from sensor-equipped devices, open up enormous amounts of information, which can be leveraged for focused attacks. Though models and algorithms have been very important in the prediction of cyber-attacks, novel methods from different data representations need to be investigated instead of task-specific techniques.

KEYWORDS: SDN Intrusion, Machine learning, Decision tree, Predicting Attack types

I. INTRODUCTION

Machine learning (ML) is a component of artificial intelligence (AI), allowing computers to learn from data without programming. Training data in ML is consumed by algorithms in making predictions based on three categorizations: supervised, unsupervised, and reinforcement learning. Classification is the major ML problem that forecasts labels for provided data, useful for speech recognition, biometric detection, and cybersecurity threat detection. Classic cybersecurity depends upon response-based systems such as firewalls and antivirus programs, usually not capable of responding to adaptive threats. Next-generation AI and ML technologies are being used today to provide proactive security measures. In this project the SDN dataset is retrieved from Kaggle website and the dataset includes 79 quantitative and qualitative features where 1 feature signifies the qualitative attributes and 78 features signify the quantitative attributes. This information will be utilized for analysis as well as to identify network intrusion. The overall information has been acquired into multiple segments that hold various kinds of network traffic.

II. OBJECTIVE

The goal of this web application based on Flask is to enable network intrusion detection through machine learning. The system enables users to upload a dataset, inspect network traffic, and classify various cyber-attacks. After uploading a dataset, the application processes the data and gives a preview of its content. It then performs analysis on the dataset to determine different types of attacks and plots a bar chart showing the distribution of attacks. For classification of network intrusions, the system utilizes several machine learning models, such as Decision Tree, Support Vector Classifier (SVC), Random Forest, and Multi-Layer Perceptron (MLP) Classifier. The dataset is divided into training and testing sets, and the performance of each model is measured in terms of accuracy and F1-score. Also, a comparison chart is created to illustrate the performance of various models. This project is intended to identify and classify cyberattacks like DDoS, XSS, Brute Force, SQL Injection, and normal traffic to help cybersecurity experts in advance threat prevention and enhancing network security using machine learning-based intrusion detection.

III. LITERATURE REVIEW

Intrusion detection in Software-Defined Networking (SDN) has become a critical area of research due to the increasing sophistication of cyber threats. Traditional security methods often struggle to keep up with SDN’s dynamic nature, making machine learning (ML) a powerful tool for identifying and preventing attacks. Researchers have explored various ML techniques to enhance intrusion detection, with models like Decision Trees, Support Vector Machines

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

(SVM), Random Forest, and Multi-Layer Perceptron (MLP) classifiers proving to be effective in analysing network traffic. For instance, Fernandes et al. (2020) demonstrated that an SVM-based intrusion detection system (IDS) could accurately detect known attacks, while Nanda et al. (2021) proposed a hybrid approach combining Decision Trees and Random Forest to improve detection rates. (2022) further expanded on this by investigating deep learning models, highlighting their potential for real-time threat detection. However, despite these advancements, challenges remain, including the need for scalable solutions, efficient feature selection, and protection against adversarial attacks. The key to strengthening SDN security lies in developing hybrid models that combine multiple techniques, improving automated feature extraction, and enhancing defense mechanism ensure robust and adaptive intrusion detection systems.

IV. SCOPE OF MY PROJECT

This project aims to develop an effective Intrusion Detection System (IDS) for Software-Defined Networking (SDN) using machine learning techniques. The goal is to accurately classify network traffic and detect cyber threats such as DDoS attacks, brute force intrusions, SQL injections, and XSS attacks, while distinguishing them from normal, benign traffic.

Key Areas of Focus

1. Dataset Processing and Analysis
- Working with a real-time network traffic dataset containing over 1.18 million observations.
- Cleaning the data by removing noise, handling missing values, and selecting the most relevant features to improve model accuracy.

2. Implementation of Machine Learning Models
- Training and testing multiple classifiers, including Decision Trees, Support Vector Machines (SVM), Random Forest, and Multilayer Perceptron (MLP).
- Evaluating model performance using key metrics such as F1-score, confusion matrix, and learning curves to ensure accurate classification.

3. Development of a Flask-Based Web Application
- Allowing users to easily upload datasets for analysis.
- Providing real-time visualizations of network traffic and classification results.
- Offering a user-friendly interface for monitoring and detecting potential threats.

4. Performance Evaluation and Optimization
- Identifying the most important features using TF-IDF and feature selection techniques.
- Conducting exploratory data analysis (EDA) to uncover patterns in attack behaviors.
- Fine-tuning models to enhance accuracy and minimize false positives.

5. Real-World Applicability
- Deploying the system in SDN environments to actively monitor and detect intrusions.
- Helping organizations strengthen their cybersecurity defenses by identifying and mitigating network threats.

V. METHODOLOGY

About the Dataset

Intrusion Detection Systems (IDS) and Prevention Systems are essential security tools that help protect network users from online threats. With the growing adoption of IoT, Cloud, and SDN technologies, networks have become more accessible and efficient. However, these advancements also introduce vulnerabilities, as cybercriminals attempt to inject malicious traffic into SDN environments to steal sensitive information. Detecting such network intrusions requires constant traffic monitoring. The dataset contains 1,188,333 observations, covering different types of network traffic, including:

- Benign traffic: 798,322 records
- DDoS attack traffic: 383,439 records
- Web attack – Brute Force: 4,550 records
- Web attack – XSS: 1,962 records
- Web attack – SQL Injection: 60 records

This dataset is used for training machine learning models to detect network intrusions effectively.

Pre processing of Data

Raw data often contains unwanted characters, symbols, and inconsistencies, making preprocessing an essential step. The data is cleaned to remove noise, missing values, and redundant information. Additionally, exploratory data analysis (EDA) is performed to check the quality of the data, ensuring that the dataset is well-structured for further analysis. This step also includes tokenization and stemming, which are commonly used text preprocessing techniques.

Feature Extraction

Feature extraction plays a crucial role in improving model performance. Using Python's scikit-learn library, various feature selection methods are applied to identify the most relevant attributes. Techniques like the bag-of-words model, n-grams, and term frequency-inverse document frequency (TF-IDF) weighting are used to refine the dataset and enhance classification accuracy.

Classification

Once the features are extracted, they are fed into different machine learning classifiers to predict cyberattacks. In this project, five classifiers from scikit-learn are used to train and evaluate the models. Each classifier is tested, and their performance is compared using F1-score and confusion matrix analysis.

The classifiers used include:

- Decision Tree Classifier: A simple yet effective model that performs multi-class classification by building a tree-like structure to make predictions. If multiple classes have the same highest probability, the model selects the one with the lowest index.
- Support Vector Classifier (SVC): This model finds the best hyperplane to separate different classes, maximizing the margin between them. Support vectors, which are the critical data points near the decision boundary, play a key role in defining the classification model.
- Random Forest Classifier: An ensemble learning method that builds multiple decision trees and combines their outputs to improve accuracy. It helps in reducing overfitting and enhances the model's robustness.
- Multilayer Perceptron (MLP) Classifier: A neural network-based model trained using backpropagation to learn complex patterns in the data, making it highly effective for classification tasks.

Exploratory Data Analysis (EDA) and Visualization

Data visualization is an essential part of understanding and analyzing data. While statistics provide numerical insights, visualizations help in gaining a qualitative understanding of patterns, anomalies, and trends in the dataset. Through various visualization techniques, key insights can be drawn, such as identifying outliers, detecting corrupted data, and understanding feature distributions. With proper domain knowledge, these visualizations can significantly enhance the interpretability of machine learning models and improve decision-making in network security.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804189

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. DATAFLOW DIAGRAM

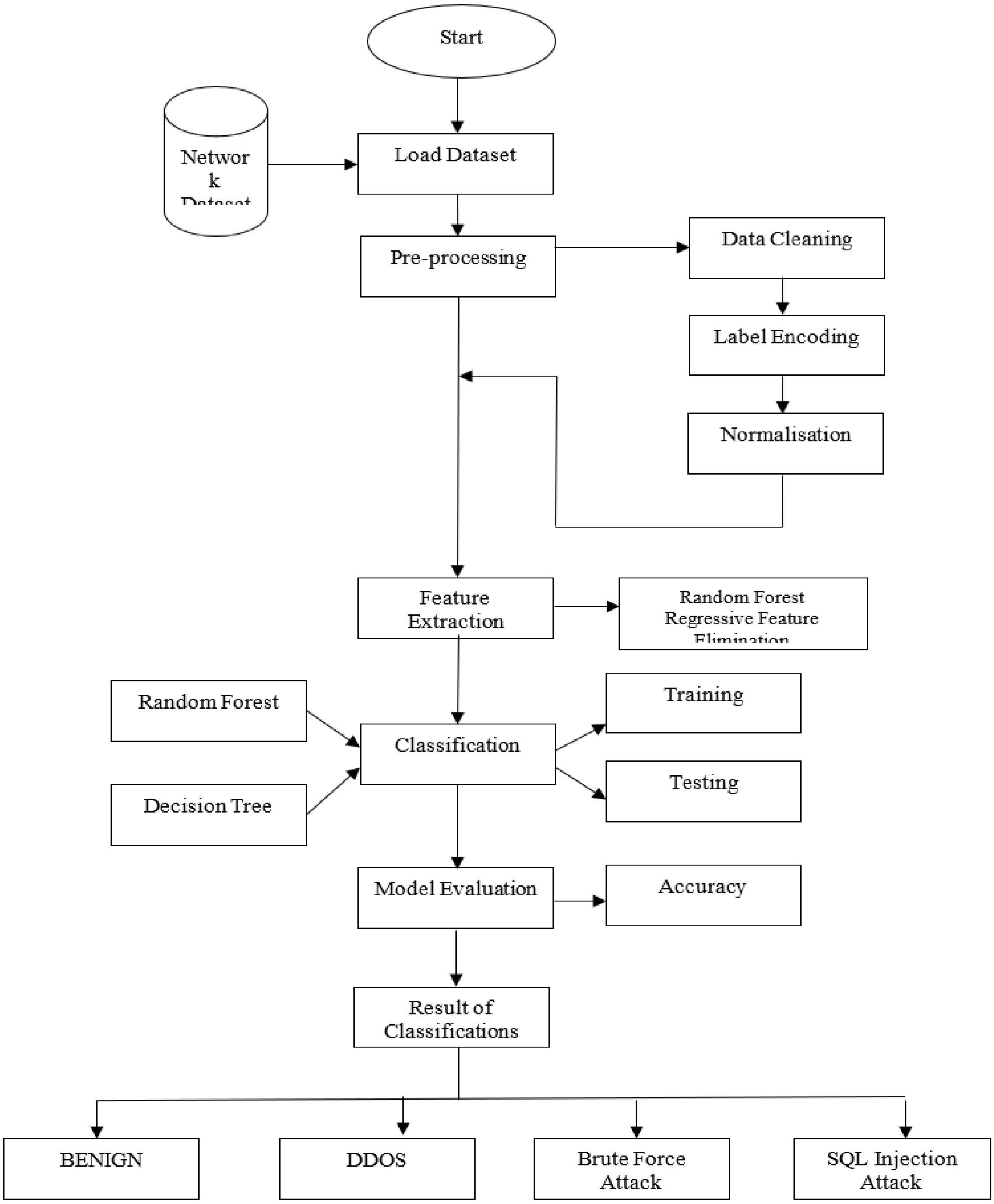


Figure 1: Data flow Diagram.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI:10.15680/IJMRSET.2025.0804189

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

SAMPLE INPUT:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
66	64	80	70975927	9	4	62	11607	20	0	6.888889	5.301991	11595	0	2901.75	5795.501	164.4079	0.183161	5914661	17700000	61300000	0	70700000	8843460
67	65	1733	3	2	0	12	0	6	6	6	0	0	0	0	0	4000000	666666.7	3	0	3	3	3	3
68	66	80	5728837	3	1	12	0	6	0	4	3.464102	0	0	0	0	2.094666	0.698222	1909612	3287200	5705320	62	5728837	2864419
69	67	52971	64	1	1	6	6	6	6	6	0	6	6	6	0	187500	31250	64	0	64	64	0	0
70	68	33608	55	1	3	0	18	0	0	0	0	6	6	6	0	327272.7	72727.27	18.33333	23.43786	45	1	0	0
71	69	49171	89477	2	1	12	6	6	6	6	0	6	6	6	0	201.169	33.52817	44738.5	63165.14	89403	74	89477	89477
72	70	38780	3	2	0	31	0	31	0	15.5	21.92031	0	0	0	0	10300000	666666.7	3	0	3	3	3	3
73	71	53	196	2	2	72	136	36	36	36	0	68	68	68	0	1061224	20408.16	65.33333	30.89229	101	47	47	47
74	72	80	68237	3	4	520	252	514	0	173.3333	295.0412	240	0	63	118.0339	11313.51	102.5836	11372.83	17185.21	33702	1	33843	16921.5
75	73	53	59971	1	1	47	175	47	47	47	0	175	175	175	0	3701.789	33.34945	59971	0	59971	59971	0	0
76	74	80	37967	3	6	26	11601	20	0	8.666667	10.2632	4380	0	1933.5	1757.79	306239.6	237.048	4745.875	9106.673	24756	17	25050	12525
77	75	56873	75592212	5	10	11613	68	11595	0	2322.6	5183.43	20	0	6.8	5.006662	154.5265	0.198433	5399444	19300000	72300000	1	75600000	18900000
78	76	57292	55	1	1	0	0	0	0	0	0	0	0	0	0	0	36363.64	55	0	55	55	0	0
79	77	50611	262	3	0	0	0	0	0	0	0	0	0	0	0	0	11450.38	131	117.3797	214	48	262	131
80	78	80	48085	3	6	26	11607	20	0	8.666667	10.2632	5755	0	1934.5	2529.125	241925.8	187.1686	6010.625	16585	47053	3	804	402
81	79	443	95125	1	1	0	0	0	0	0	0	0	0	0	0	0	21.02497	95125	0	95125	95125	0	0
82	80	443	6612939	16	16	2028	12256	850	0	126.75	234.6016	2920	2	766	832.3309	2160.008	4.838998	213320.6	970863.7	5414110	1	6612939	440862.6
83	81	443	186158	47	56	1319	108737	570	0	28.06383	110.2995	2896	0	1941.732	922.2032	591196.7	553.2934	1825.078	6693.358	44310	1	186158	4046.913
84	82	80	1757448	3	6	26	11607	20	0	8.666667	10.2632	5755	0	1934.5	2529.125	6619.257	5.121062	219681	621001.8	1756582	4	580	290
85	83	443	1.19E+08	40	39	31166	20855	5792	0	779.15	1630.787	1461	0	534.7436	657.4077	438.109	0.66532	1522306	3514908	10000000	1	1.19E+08	3044612
86	84	53	170	2	2	90	122	45	45	45	0	61	61	61	0	1247059	23529.41	56.66667	57.88206	118	3	3	3
87	85	443	313	2	0	37	0	37	0	18.5	26.16295	0	0	0	0	118210.9	6389.776	313	0	313	313	313	313
88	86	443	5664045	8	5	372	4993	191	0	46.5	71.87688	1815	0	998.6	812.0929	947.2029	2.29518	472003.8	1535437	5346309	27	5664045	809149.3
89	87	80	855036	3	5	26	11607	20	0	8.666667	10.2632	5840	0	2321.4	3173.374	13605.28	9.356331	122148	322828.9	854255	41	516	258
90	88	80	6978	2	0	0	0	0	0	0	0	0	0	0	0	0	286.6151	6978	0	6978	6978	6978	6978
91	89	80	1622031	4	0	24	0	6	6	6	0	0	0	0	0	14.79626	2.466044	540677	931287.8	1616030	1	1622031	540677
92	90	80	1.17E+08	18	15	1298	506	604	0	72.11111	193.5081	250	0	33.73333	87.81756	15.48314	0.283228	3641058	4772827	10000000	25	1.17E+08	6853756
93	91	42248	4	3	0	77	0	46	0	25.66667	23.45918	0	0	0	0	19300000	750000	2	1.414214	3	1	4	2
94	92	443	71666	9	6	567	6368	208	0	63	80.8146	2636	0	1061.333	1010.227	96768.34	209.3043	5119	8156.365	22625	32	71666	8958.25
95	93	80	5197465	3	1	0	0	0	0	0	0	0	0	0	0	0	0.769606	1732488	3000158	5196772	134	5197465	2598733
96	94	443	5289883	9	5	348	3763	191	0	38.66667	70.5886	1448	0	752.6	670.6003	777.1438	2.646561	406914.1	1439194	5196686	2	5289883	661235.4

Figure 2: SDN dataset

VII. RESULT

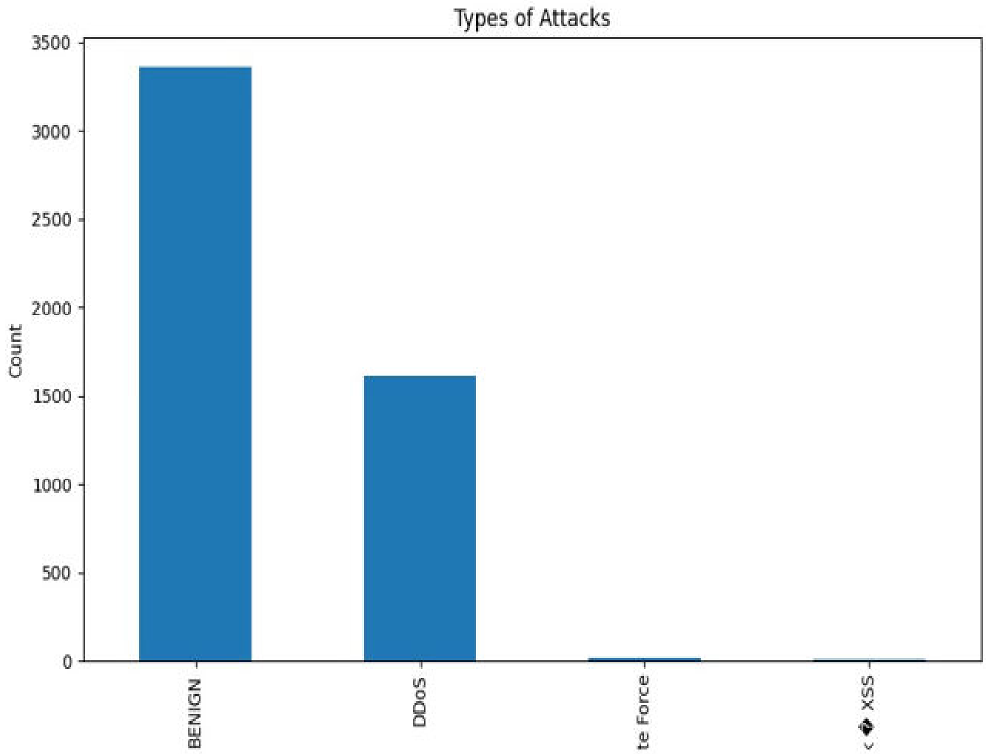


Figure 3: Prediction result

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

DecisionTree	0.996	0.9959398182064848
SVC	0.872	0.8675644698958153
RandomForest	0.9966666666666667	0.9963134328156245
MLPClassifier	0.968	0.9655387373292916

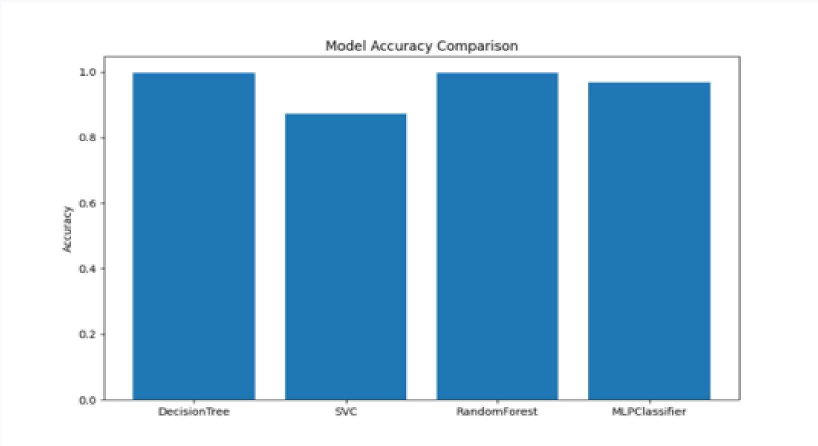


Figure 4: Model Accuracy prediction

VIII. CONCLUSION

This project presents an effective approach to intrusion detection in Software-Defined Networking (SDN) using machine learning techniques. By leveraging a dataset containing real-time network traffic, the system classifies various types of cyber threats, such as DDoS attacks, brute force intrusions, SQL injections, and XSS attacks, while distinguishing them from benign traffic. The implementation of multiple classifiers—Decision Trees, Support Vector Machines (SVM), Random Forest, and Multilayer Perceptron (MLP)—ensures a comprehensive evaluation of different machine learning models. Additionally, the integration of a Flask-based web application enhances usability by providing real-time data visualization and an intuitive interface for network monitoring.

FUTURE ENHANCEMENT

Incorporating deep learning models to improve detection accuracy.

- Expanding the dataset to include a wider range of attack types and real-world scenarios.
- Developing an adaptive system that continuously learns from new threats to improve security over time.

By combining machine learning with real-time network analysis, this project lays the groundwork for an intelligent, scalable, and adaptive IDS that enhances SDN security.

REFERENCES

1. Fernandes, D., Rodrigues, J. J. P. C., Carvalho, L. F., Al-Muhtadi, J., & Proença, M. L. (2020). "A comprehensive survey on network anomaly detection." Telecommunication Systems, 73(3), 447–489.
2. Nanda, A., Puthal, D., Mohanty, S. P., Prasad, M., & Liu, Y. (2021). "Machine learning-based intrusion detection for software-defined networking." IEEE Transactions on Network and Service Management, 18(2), 1201-1215.
3. Alshamrani, A., Myneni, S., Chowdhary, A., & Huang, D. (2022). "A deep learning-based intrusion detection system for SDN-enabled networks." IEEE Access, 10, 1532-1544.
4. Scikit-learn. (2023). "Machine Learning in Python." Retrieved from https://scikit-learn.org
5. These references provide valuable insights into the methodologies and technologies used in this project, supporting its scientific and technical foundation.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com