# INTELLIGENT CYBER ATTACK DETECTION SYSTEM FOR IOT NETWORKS USING MACHINE LEARNING TECHNIQUES

*Report submitted to SASTRA Deemed to be University as the requirement for the course*

# ICT300 - MINI PROJECT

*Submitted by*

## SRIVIDYA.N

**(Reg.No: 124014092, B.Tech-ICT) THARUNYA .S (Reg.No:124014056, B.Tech-ICT) DHARINI.S.N (Reg.No:124014083, B.Tech-ICT)**

## May 2023



SCHOOL OF COMPUTING

**THANJAVUR, TAMIL NADU, INDIA - 613 401**



SCHOOL OF COMPUTING

**THANJAVUR-613001**

# Bonafide Certificate

This is to certify that the report titled “**Intelligent Cyber Attack Detection System for IOT Networks using Machine Learning Techniques**” submitted as a requirement for the course, ICT300: **MINI PROJECT** for B.Tech. is a bonafide record of the work done by **Ms. Tharunya S(Reg.No:124014056, B.Tech-ICT), Ms. Dharani S.N(Reg.No:124014083, B.Tech-ICT), Ms.**

**Srividya N(Reg.No: 124014092, B.Tech-ICT)** during the academic year 2022-23, in the School of Computing, under my supervision.



## Signature of Project Supervisor :

**Name with Affiliation :** Dr. Hariharan K, Asst.Professor, ICT

## Date : 15-05-2023

Mini Project *Viva voc*e held on

**Examiner 1 Examiner 2**

# ACKNOWLEDGEMENTS

We would like to thank our Honorable Chancellor **Prof. R. Sethuraman** for providing us with an opportunity and the necessary infrastructure for carrying out this project as a part of our curriculum.

We would like to thank our Honorable Vice-Chancellor **Dr. S. Vaidhyasubramaniam** and **Dr. S. Swaminathan,** Dean**,** Planning & Development, for the encouragement and strategic support at every step of our college life.

We extend our sincere thanks to **Dr. R. Chandramouli,** Registrar, SASTRA Deemed to be University for providing the opportunity to pursue this project.

We extend our heartfelt thanks to **Dr. A. Umamakeswari**, Dean, School of Computing, **Dr. B.Santhi**, Associate Dean, Department of Computer Application, **Dr. R. Muthaiah**, Associate Dean, Department of Information Technology and Information & Communication Technology and **Dr. V. S. Shankar Sriram**, Associate Dean, Department of Computer Science and Engineering for their motivation and support offered in materializing this project.

Our guide **Dr. Hariharan K, Asst.Professor**, School of Computing was the driving force behind this whole idea from the start. His deep insight in the field and invaluable suggestions helped us in making progress throughout our project work. We also thank the members of the project review panel for their insightful remarks and recommendations that improved our project.

We’d like to express our gratitude to all of the teaching and non-teaching faculty members at the School of Computing who have assisted us in finishing the project in any way.

We gratefully acknowledge all the contributions and encouragement from my family and friends resulting in the successful completion of this project. We thank you all for providing me an opportunity to showcase my skills through projects.

# LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Title** | **Page No** |
| 4.1 | Pearson correlation coefficient heatmap | 10 |
| 4.2 | Random forest mean decrease accuracy plot | 11 |
| 4.3 | Information gain ratio plot | 11 |
| 4.4 | Reduced final feature set | 12 |
| 4.5 | Comparative chart of accuracy | 12 |
| 4.6 | Comparative plot for training time | 13 |
| 4.7 | Bar chart for overfitting | 13 |

**ABBREVIATIONS**

IDS - Intrusion Detection System IoT - Internet of Things

DoS - Denial of Service

DS2OS - Distributed Smart Space Orchestration System PCC - Pearson Correlation Coefficient

GR - Gain Ratio

RFMDA - Random Forest Mean Decrease Accuracy AC - Accuracy

# Abstract

Intrusion detection system (IDS) has a significant role in the detection and prevention of cyberattacks in IoT networks. Yet, the dimensionality curse in the majority of current IDS lowers the effectiveness of IoT systems as a whole. So, it's crucial to cut out redundant and pointless elements while creating effective IDS. The present research proposes an innovative hybrid feature reduction strategy for an intelligent cyber-attack detection system for IoT networks. To get three separate feature sets, this approach performs ranking of features using random forest mean decrease accuracy, correlation coefficient and gain ratio. Then, using a well-developed method (AND operation), features are joined resulting in a single ideal feature set. The newly reduced feature set is given to six prominent machine learning algorithms. Using the most recent DS2OS dataset, the effectiveness of the introduced cyber-attack detection algorithm is evaluated. In terms of accuracy, training duration, and overfitting, the proposed work is studied and contrasted with various recent state-of-the-art methods.

**KEY WORDS:** Intrusion Detection System, Dimensionality Curse, Feature Reduction, Random Forest Mean Decrease Accuracy, Information Gain Ratio, Pearson Correlation Coefficient.

# Table of Contents

## Title Page No.

[Bonafide Certificate i](#_bookmark0)

[Acknowledgements ii](#_bookmark1)

[List of Figures iii](#_bookmark2)

Abbreviations iv

[Abstract v](#_bookmark3)

1. Summary 1
2. Merits and Demerits of the base paper 3
3. Source Code 4
4. Snapshots 10
5. Conclusion and Future Plans 14
6. References 15
7. Appendix 16

# CHAPTER 1 SUMMARY OF THE BASE PAPER

**Summary:**

**Base Paper Details:**

Title : Toward Design of an Intelligent Cyber Attack Detection System using Hybrid Feature Reduced Approach for IoT Networks

Journal : Arabian Journal for Science and Engineering

Author : Prabhat kumar, Govindh P.Gupta, Rakesh Tripathi

Publisher: Received 03 June 2020; Accepted 24 November 2020; published on 11 January 2020.

Citation Index: Web of Science

**Introduction:** The base paper introduces a novel approach for identifying cyber-attacks in IoT networks. The paper starts by highlighting the growing importance of IoT networks and the potential security threats that they face. The authors then propose a hybrid feature reduction approach that combines various machine learning techniques to detect cyber-attacks in real-time.

**Novelty:** The authors suggest a hybrid feature reduction approach that combines wrapper and filter-based feature selection techniques to generate an intrusion detection system feature set that is optimal.

**Algorithm:** This hybrid model incorporates gain ratio, mean decrease accuracy, and correlation coefficient. The three subsets of features that are obtained as a result of these methods are then ranked according to their respective feature scores. Applying the intersection technique (AND operation) results in the most important feature set.

# Dataset

**Dataset Name**: DS2OS traffic traces

**Description:** This database contains 359752 instances and 12 attributes. The presence or absence of the target variable of a cyber-attack, which is to be predicted based on a set of input features. DS2OS (Distributed smart space orchestration system) is a research project that aims to develop a platform for managing smart spaces. The DS20S traffic traces dataset consists of a collection of network traffic traces captured from a simulated smart city environment. The dataset was captured using a traffic emulator tool (TET), which generates realistic network traffic by simulating various types of IoT devices and their interactions with a smart city network.

# CHAPTER 2

**MERITS AND DEMERITS OF THE BASE PAPER**

## Merits:

* The proposed approach addresses the growing security threats faced by IoT networks, which is an important area of research.
* The use of a hybrid feature reduction approach that combines multiple machine learning techniques allows for accurate and efficient detection of cyber-attacks.
* Its usefulness is demonstrated by the examination and comparison of the suggested approach utilizing different datasets.
* The approach can be applied in real-time, which is crucial for securing the IoT networks.
* The report offers insightful information on the application of sophisticated machine learning methods for cyber threat detection.

## Demerits:

* The proposed approach may require significant computational resources and expertise to implement, which could limit its applicability in certain settings.
* The paper does not provide a detailed discussion of the limitations or potential challenges of the proposed approach.
* The datasets used for evaluation may not fully represent real- world scenarios, which could limit the generalizability of the result
* The paper does not discuss the scalability of the proposed approach for larger IoT networks.
* The proposed approach does not address potential issues related to privacy or data protection in IoT networks.

# CHAPTER 3 SOURCE CODE

**FEATURE SELECTION:**

**# Pearson Correlation Coefficient**

from scipy.stats import pearsonr pcc = []

for i in range(X\_scaled.shape[1]): corr\_all, \_ = pearsonr(X\_scaled[:, i], y) pcc.append(abs(corr\_all))

pcc = np.array(pcc)

sorted\_idx = pcc.argsort()[::-1]

# Print PCC value for each feature in descending order for i in sorted\_idx:

print(f"{X.columns[i]}: {pcc[i]}")

# Select top k features k = 9

top\_k\_features\_pcc = X.columns[sorted\_idx[:k]].values print("\nTop {} features: {}".format(k, top\_k\_features\_pcc))

## #Random Forest

from sklearn.ensemble import RandomForestClassifier # Train a Random Forest classifier on the dataset

rf = RandomForestClassifier(n\_estimators=100, random\_state=42) rf.fit(X\_scaled, y)

# Compute feature importance scores importances = rf.feature\_importances\_

# Sort features by importance score in descending order indices = np.argsort(importances)[::-1]

# Print feature importance score for each feature in descending order for i in indices:

print(f"{X.columns[i]}: {importances[i]}")

# Select top k features k = 9

top\_k\_features\_rfmda = X.columns[indices[:k]].values

print("Top {} features: {}".format(k, top\_k\_features\_rfmda)) from sklearn.feature\_selection import mutual\_info\_classif **#Information Gain Ratio**

from sklearn.feature\_selection import SelectKBest

# Apply SelectKBest to select the top k features using mutual information gain ratio k = 9

selector = SelectKBest(mutual\_info\_classif, k=k) selector.fit(X\_scaled, y)

# Get the selected feature indices

selected\_features = selector.get\_support(indices=True)

# Get the feature scores scores = selector.scores\_

# Sort features by score in descending order indices = np.argsort(scores)[::-1]

# Print feature score for each feature in descending order for i in indices:

print(f"{X.columns[i]}: {scores[i]}") # Select top k features

top\_k\_features\_gr = X.columns[selected\_features].values print("Top {} features: {}".format(k, top\_k\_features\_gr))

## #Final feature set:

final\_features = list(set(top\_k\_features\_gr) & set(top\_k\_features\_rfmda) & set(top\_k\_features\_pcc))

print("Final feature set:", final\_features)

## ALGORITHMS:

**KNN:**

knn = KNeighborsClassifier(n\_neighbors=5) knn.fit(X\_train1, y\_train1)

**RANDOM FOREST:**

rf = RandomForestClassifier(n\_estimators=50, random\_state=42) rf.fit(X\_train1, y\_train1)

y\_pred = rf.predict(X\_test1)

**LOGISTIC REGRESSION:**

logreg = LogisticRegression() logreg.fit(X\_train1, y\_train1) y\_pred = logreg.predict(X\_test1)

## DECISION TREE:

from sklearn.tree import DecisionTreeClassifier dtc = DecisionTreeClassifier()

dtc.fit(X\_train1, y\_train1) y\_pred = dtc.predict(X\_test1)

## NAIVE BAYES:

from sklearn.naive\_bayes import GaussianNB nb = GaussianNB()

nb.fit(X\_train1, y\_train1) y\_pred = nb.predict(X\_test1)

## XGBOOST:

import xgboost as xg

train\_data = xg.DMatrix(X\_train1, label=y\_train1) test\_data = xg.DMatrix(X\_test1, label=y\_test1)

params = { 'max\_depth': 3,

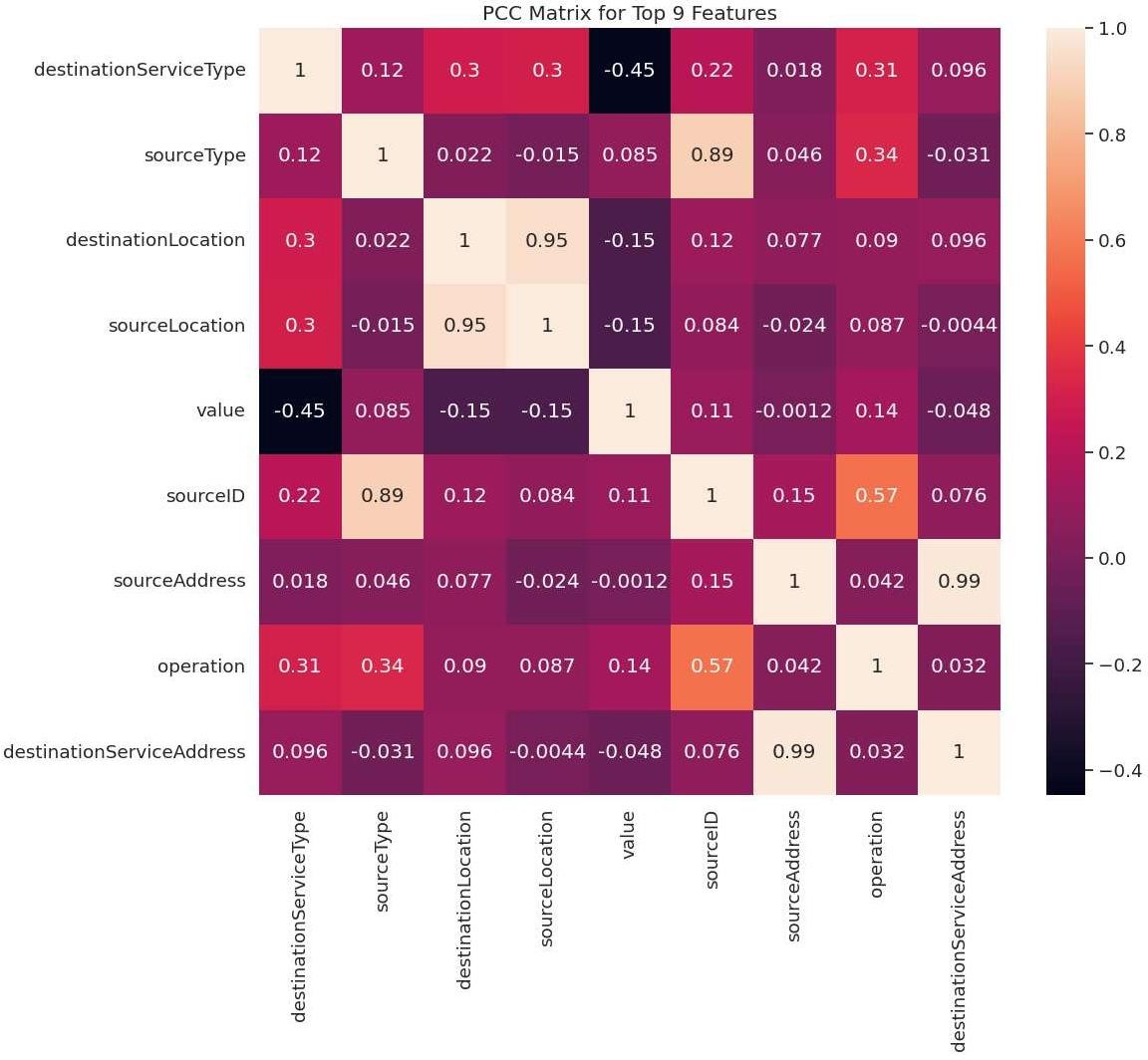
'eta': 0.1,

'objective': 'multi:softmax', 'num\_class': 8

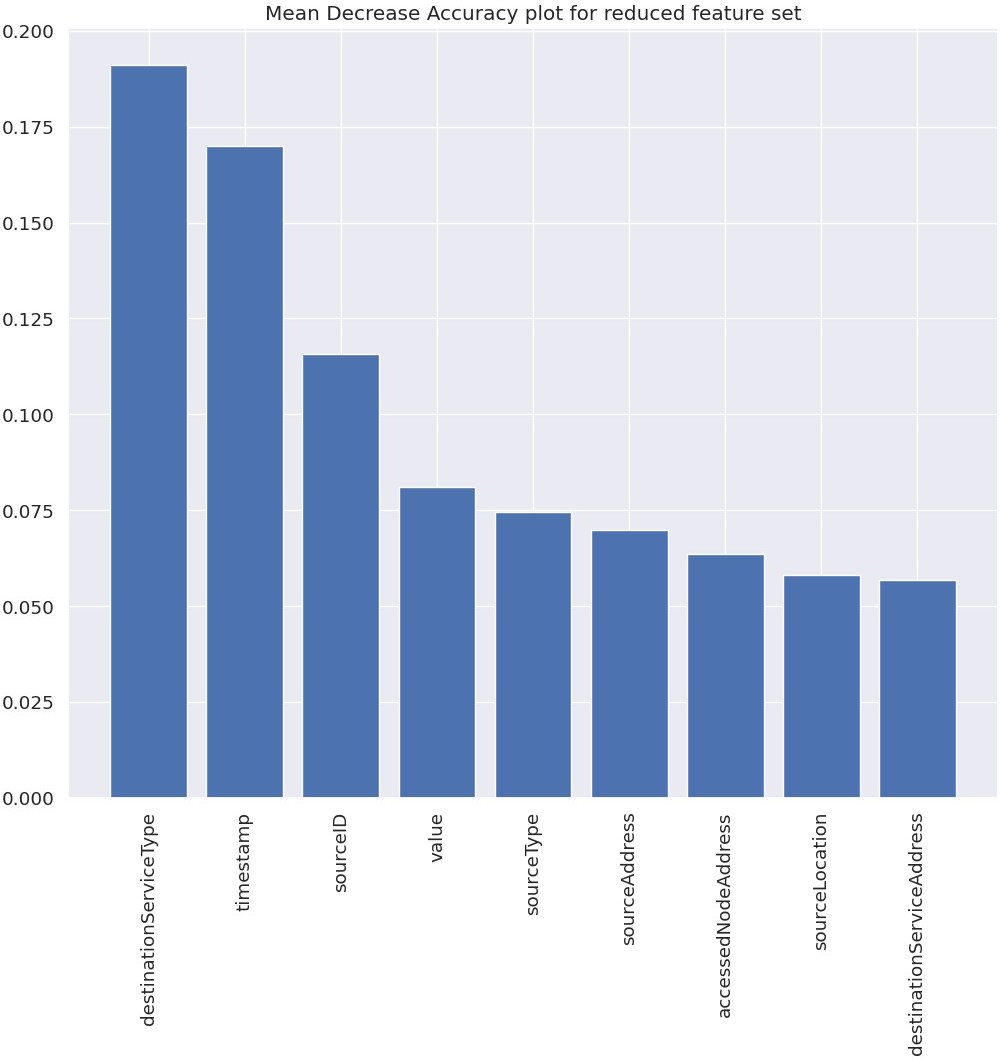
}

model = xg.train(params, train\_data, num\_boost\_round=10) predictions = model.predict(test\_data)

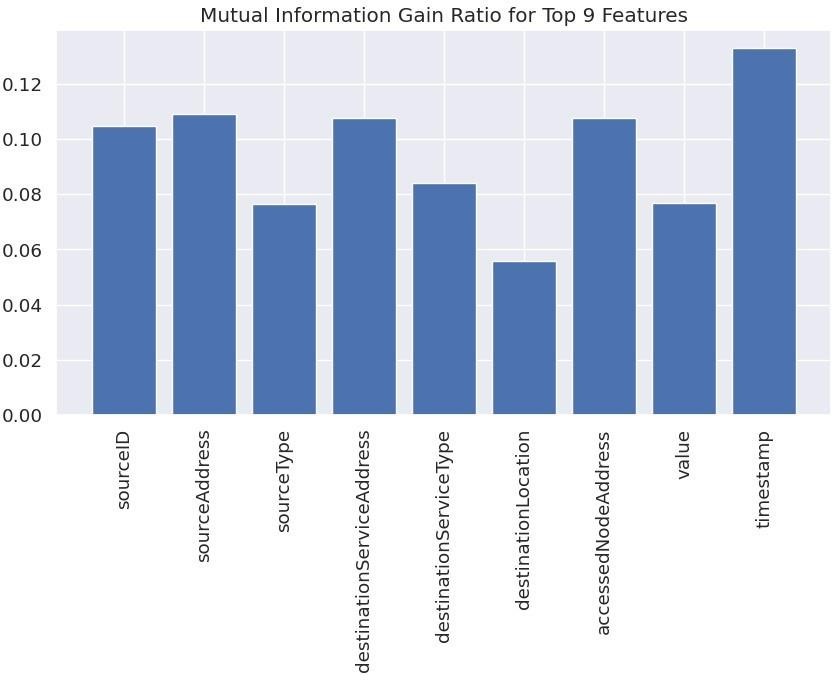
# CHAPTER 4 SNAPSHOTS



**fig 4.1**



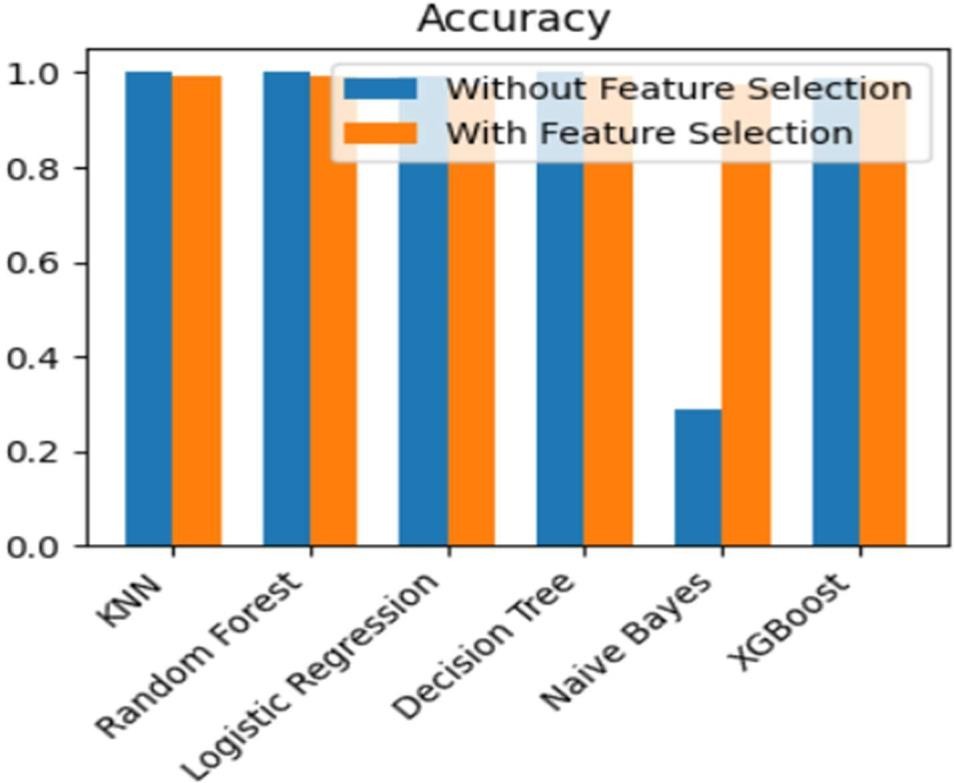
# fig 4.2



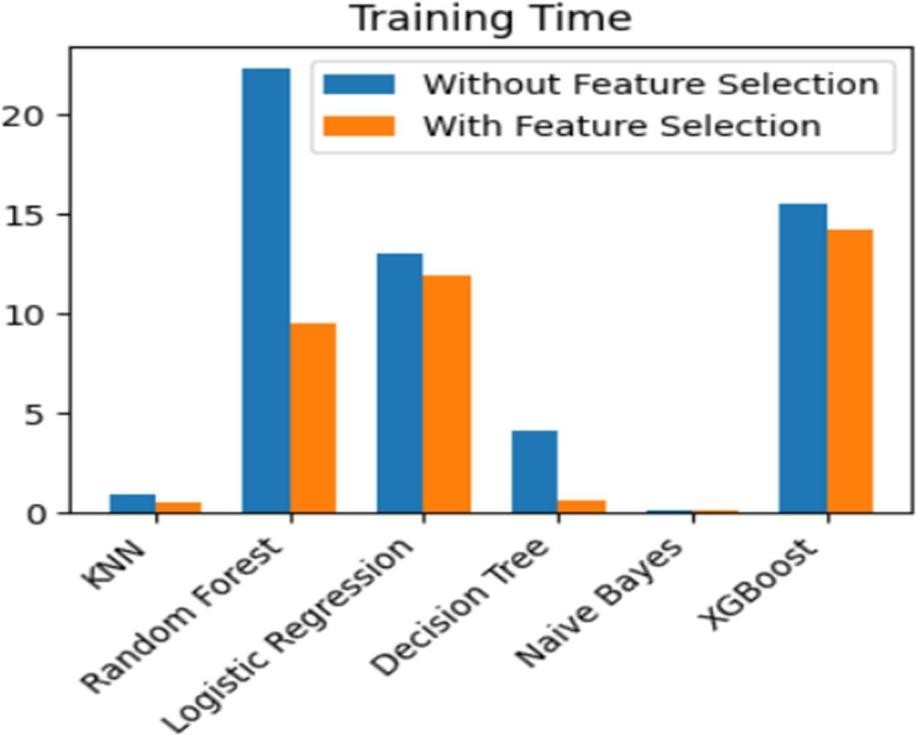
**fig 4.3**



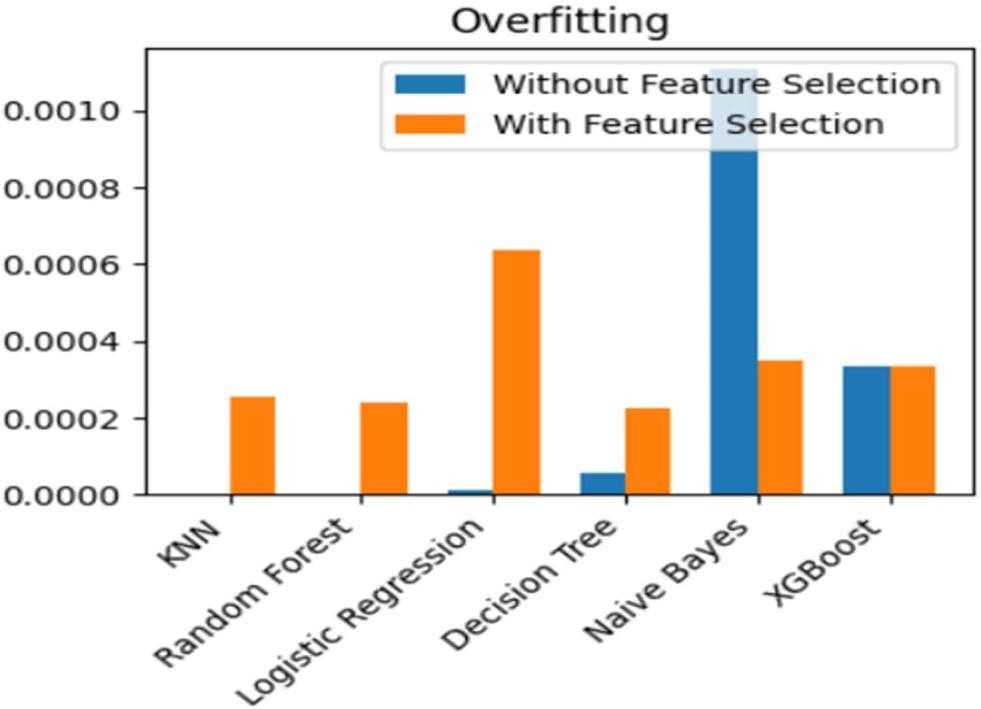
# fig 4.4



**fig 4.5**



# fig 4.6



**fig 4.7**

# CHAPTER 5 CONCLUSION AND FUTURE PLANS

**Conclusion:** A significant advancement in the area of cyber security is the development of a system for detecting intelligent cyberattacks for Internet of Things networks applying machine learning techniques. The experiment has shown that machine learning techniques are highly accurate and efficient at identifying cyberattacks on IoT networks. Since numerous devices generate a lot of data, an important step in the creation of an IoT detection system is feature selection since it reduces the dimension of the data, increases the system's accuracy, and aids in the detection of IoT attacks.

Overall, this project shows the potential for machine learning to improve cybersecurity and shield organizations from the growing hazards of cyber-attacks in IoT networks.

# Future Plans:

1. Hybrid Detection Systems: Combine multiple detection techniques.
2. Reinforcement Learning: Reinforcement learning algorithms can learn from feedback and adapt to new threats in real-time, making them highly effective in dynamic environments like IoT networks.
3. Improved Feature Selection
4. Integration with Blockchain Technology: Future research could explore the integration of blockchain technology with the detection system to provide a secure and transparent mechanism for storing and sharing data related to cyber-attacks in IoT networks.

**CHAPTER 6**

**REFERENCES:**

1. Tama, B.A.; Comuzzi, M.; Rhee, K.-H.: TSE-IDS: a two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. IEEE Access 7, 94 497–94 507 (2019)
2. Gao, X.; Shan, C.; Hu, C.; Niu, Z.; Liu, Z.: An adaptive ensemble machine learning model for intrusion detection. IEEE Access 7, 82 512–82 521 (2019)
3. Latah, M.; Toker, L.: Towards an efficient anomaly-based intrusion detection for software-defined networks. IET Netw. 7(6), 453–459 (2018)
4. Staudemeyer, R.; Omlin, C.: Feature set reduction for automatic network intrusion detection with machine learning algorithms. In: Proceedings of the Southern African Telecommunication Networks and Applications Conference (SATNAC), p. 105 (2009)
5. Latah, M.; Toker, L.: An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks. CoRR (2018)

**CHAPTER 7**

**APPENDIX**

Base Paper Link: <https://link.springer.com/article/10.1007/s13369-020-05181-3>