iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** V **Month of publication:** May 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.42311

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 10 Issue V May 2022- Available at www.ijraset.com*

# An Efficient Spam Detection Technique for IoT Devices Using Machine Learning

Mr. G. Sekhar Reddy[1], N. Rakesh Naik[1], A. Prashanth[3], B. Pranay Kumar[4]
[1, 2]*Assistant Professor, Dept of IT, Anurag Group of Institutions, Telangana*
[1, 2, 3, 4]*Bachelor of Technology, Information Technology, Anurag Group of Institutions, Telangana, India*

***Abstract:*** ***The Internet of Things (IoT) is a group of millions of devices having sensors and actuators linked over wired or wireless channel for data transmission. IoT has grown rapidly over the past decade with more than 25 billion devices are expected to be connected by 2020. The volume of data released from these devices will increase many-fold in the years to come. In addition to an increased volume, the IoT devices produces a large amount of data with a number of different modalities having varying data quality defined by its speed in terms of time and position dependency. In such an environment, machine learning algorithms can play an important role in ensuring security and authorization based on biotechnology, anomalous detection to improve the usability and security of IoT systems. On the other hand, attackers often view learning algorithms to exploit the vulnerabilities in smart IoT-based systems. Motivated from these, in this paper, we propose the security of the IoT devices by detecting spam using machine learning.***

***Keywords: SVM - Support Vector Machine , KNN - K-NN Algorithm , NNM - Neural Network Model, RFE - Recursive Feature Elemination.***

## I. INTRODUCTION

IoT is taken into account as an interconnected and distributed network of embedded systems communicating through wired or wireless communication technologies. Massive growth and rapid development in the field of the Internet of Things (IoT), makes the presence of IoT devices prevalent in smart homes and smart cities. It is also defined because the network of physical objects or things empowered with limited computation, storage, and communication capabilities is also embedded with electronics (such as sensors and actuators), software, and network connectivity that permits these objects to gather, sometimes process, and exchange data. The things in IoT ask the objects from our lifestyle starting from smart household devices like a smart bulb, smart adapter, smart meter, smart refrigerator, smart oven, AC, temperature sensor, smoke detector, IP camera, to more sophisticated devices like frequency Identification (RFID) devices, heartbeat detectors, accelerometers, sensors in the parking zone, and a variety of other sensors in automobiles, etc.

There are various large amounts of applications and services offered by the IoT ranging from critical infrastructure to agriculture, military, home appliances, and personal health care. As the usage of IoT devices increases the anomalies generated by these devices also grow beyond the count. IoT applications need to ensure information protection to fix security issues like interruptions, spoofing attacks, Dos attacks, jamming, eavesdropping, spam, and malware. The maximum care to be taken is with web-based devices as the maximum number of IoT devices are web-dependent. It is common in the work environment that the IoT devices introduced in an association can be utilized to execute security and protection includes proficiently. For example, wearable devices that collect and send user's health data to a connected smartphone should prevent leakage of data to ensure privacy. It has been found in the market that 25-30% of working employees connect their Personal IoT devices with the organizational network. The expanding nature of IoT attracts both the audience, i.e., the users and therefore the attackers.

However, with the emergence of ML in various attack scenarios, IoT devices choose a defensive strategy and decide the key parameters in the security protocols for a trade-off between security, privacy, and computation. This work enhances the algorithm to affect the time-series regression model rather than a classification model and may also execute ML models in parallel. This proposed paper focuses on determining the trustworthiness of the IoT device within the smart home network. The algorithm scores an IoT device with a spamicity score to secure smart devices by calculating spam scores using different machine learning models.

## II. LITERATURE SURVEY

It is often used by computer attackers to characterize hosts or networks which they are considering hostile activity against. Thus it is useful for system administrators and other network defenders to detect portscans as possible preliminaries to a more serious attack. It is also widely used by network defenders to understand and find vulnerabilities in their own networks.

### *A. Scope- Based*

As part of the recommended approach, the spammy characteristics are detected. ML models are used in Internet of things. This is the IoT data. it is pre-processed with the aid of pattern development method. By playing around with the structure, each IoT device is rewarded with ML models. T he amount of spamming that has been detected As a result, the criteria for success have been refined. IoT equipment operating in a smart house As we go forward, will take into account meteorological conditions as well as the environment IoT devices more secured and reliable.

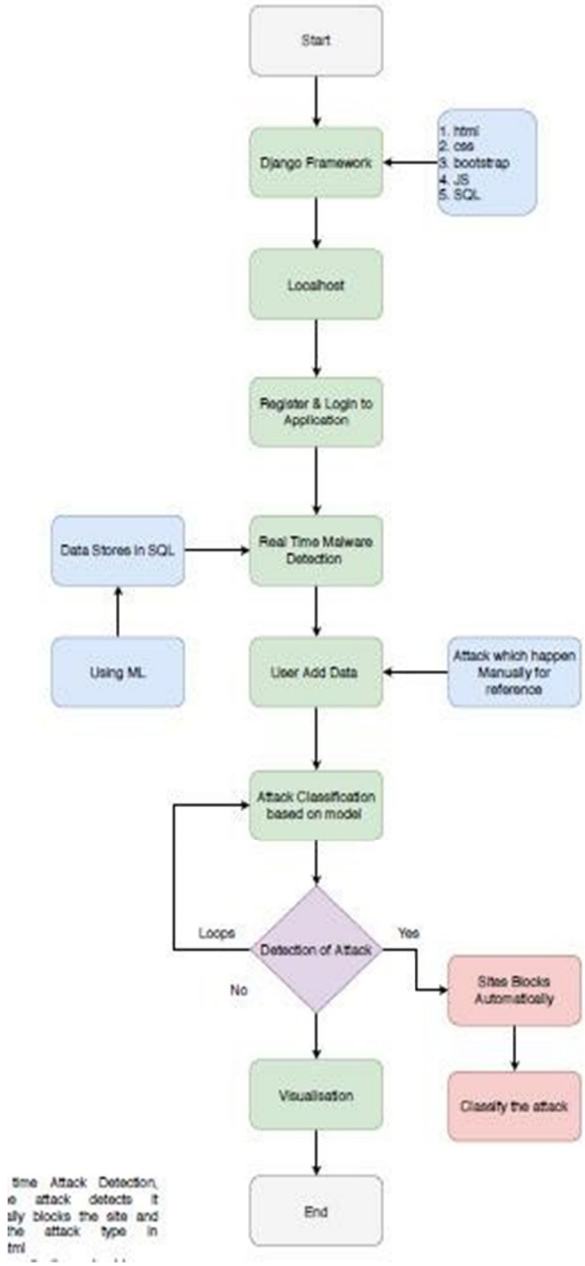



Fig 2.1: Flow Diagram of an efficient spam detection technique for iot devices using machine learning

*B. Motivation*

The use of new innovations give incredible advantages to people, organizations, and governments, be that as it may, messes some up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information and so forth. Contingent upon these issues, digital fear based oppression is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and nation security by different gatherings, for example, criminal association, proficient people and digital activists. Along these lines, Intrusion Detection Systems (IDS) has been created to maintain a strategic distance from digital assaults.

## III. EXISTING SYSTEM

Blameless Bayes and Principal Component Analysis (PCA) were been used with the KDD99 dataset by Almansob and Lomte [9].Similarly, PCA, SVM, and KDD99 were used Chithik and Rabbani for IDS [10]. In Aljawarneh et al's. Paper, their assessment and examinations were conveyed reliant on the NSL-KDD dataset for their IDS model [11] Composing inspects show that KDD99 dataset is continually used for IDS [6]–[10].There are 41 highlights in KDD99 and it was created in 1999. Consequently, KDD99 is old and doesn't give any data about cutting edge new assault types, example, multi day misuses and so forth. In this manner we utilized a cutting-edge and new CICIDS2017 dataset [12] in our investigation.

## IV. PROPOSED SYSTEM

Important steps of the algorithm are given in below. 1) Normalization of every dataset. 2) Convert that dataset into the testing and training. 3) Form IDS models with the help of using RF, ANN, CNN and SVM algorithms. 4) Evaluate every model's performances

## V. ARCHITECTURE

The architecture provides the entire process flow of the system.

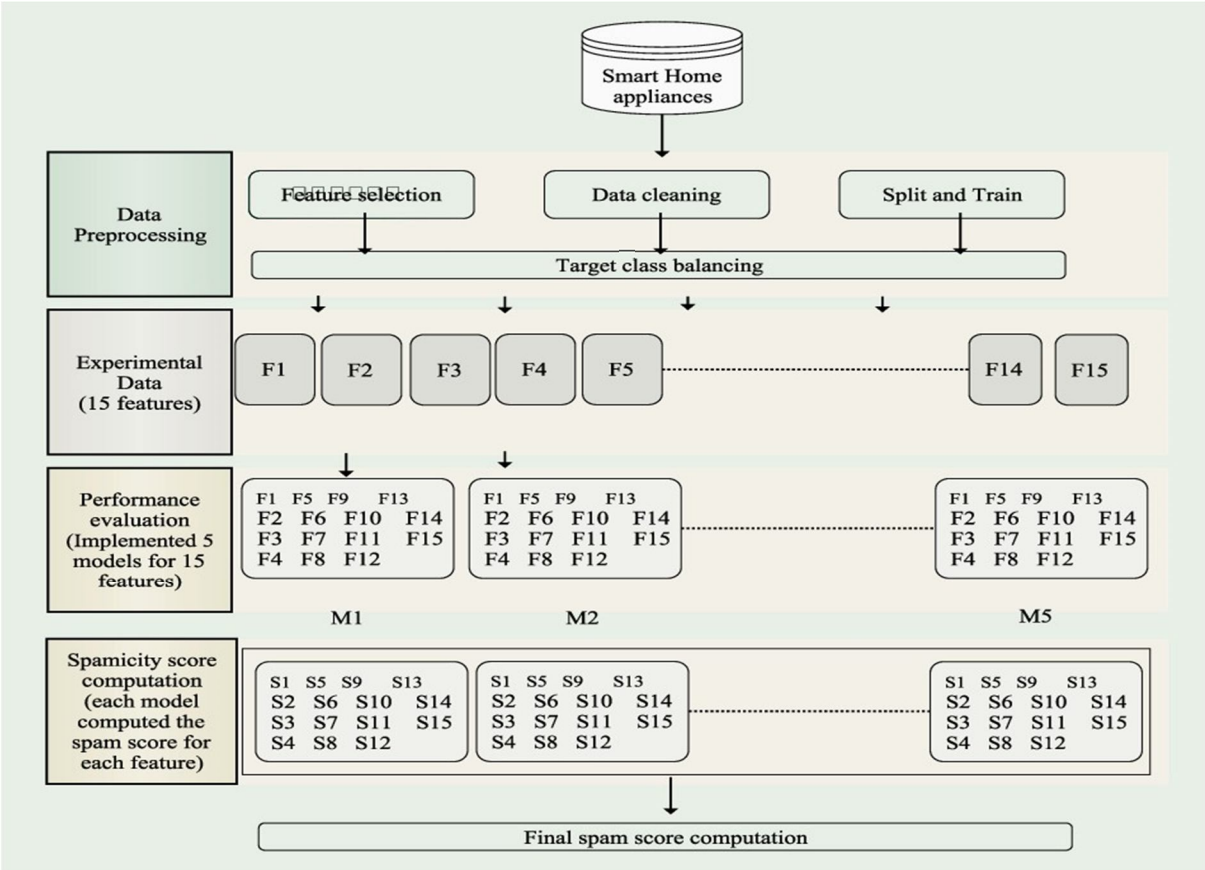



Fig-5 Architecture diagram

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue V May 2022- Available at www.ijraset.com*

## VI. IMPLEMENTATION

### A. *Algorithms Used*

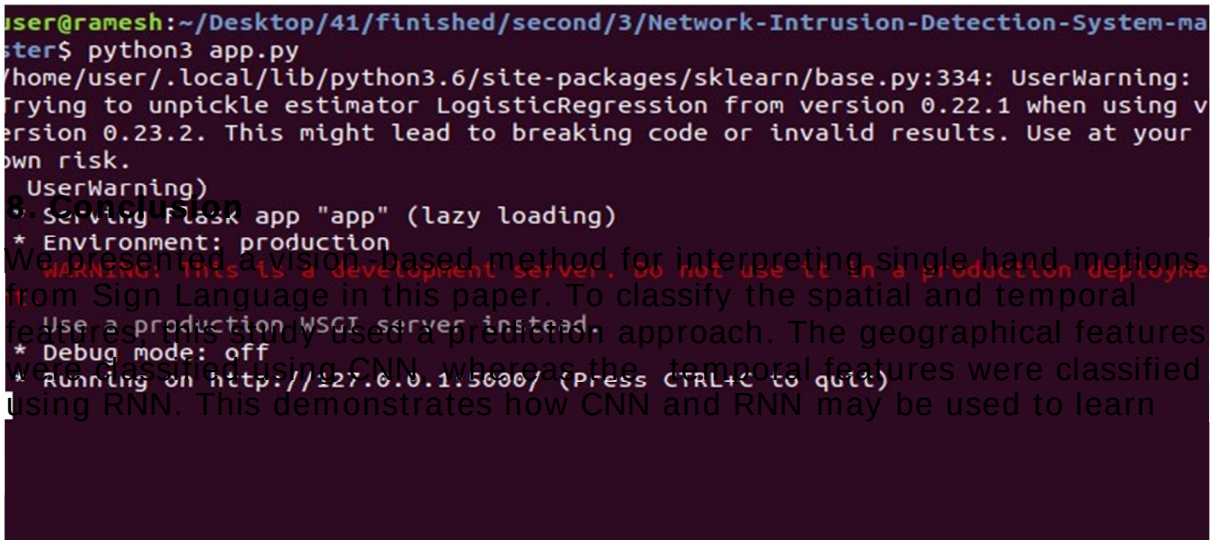
1) *Decision Tree:* The goal of this algorithm is to create a model that predicts the value of a target variable, for which the decision tree uses the tree representation to solve the problem in which the leaf node corresponds to a class label and attributes are represented on the internal node of the tree.
2) *Logistic Regression:* Logistic regression uses the concept of predictive modeling as regression; therefore, it is called logistic regression, but is used to classify samples; Therefore, it falls under the classification algorithm.
3) *Support Vector Machine:* Support Vector Machine(SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems.
4) *K-Nearest Neighbour:* K-Nearest Neighbour. It is a supervised machine learning algorithm. The algorithm can be used to solve both classification and regression problem statements. The number of nearest neighbours to a new unknown variable that has to be predicted or classified.
5) *Recursive Feature Elimination:* Recursive Feature Elimination, or RFE for short, is a popular feature selection algorithm. RFE is popular because it is easy to configure and use and because it is effective at selecting those features (columns) in a training dataset that are more or most relevant in predicting the target variable.
6) *Neural Network Model:* An artificial neural network learning algorithm, or neural network, or just neural net. , is a computational learning system that uses a network of functions to understand and translate a data input of one form into a desired output, usually in another form.

### B. *Methodology*

Testing is a process of executing a program with the aim off inding error. To make our software perform well it should be error free. If testing is done successfully it will remove all the errors from the software.

1) *Unit Testing:* Software verification and validation method in which a programmer tests if individual unitsof source code are fit for use. It is usually conducted by the development team.
2) *Integration Testing:* The phase in software testing in which individual software modules are combined and tested as a group. It is usually conducted by testing teams.
3) *Alpha Testing:* Type of testing a software product or system conducted at the developer's site. Usually it is performed by the end users.
4) *Beta Testing:* Final testing before releasing application for commercial purpose. It is typically done by end- users or others.
5) *Performance Testing:* Functional testing conducted to evaluate the compliance of a system or component with specified performance requirements. It is usually conducted by the performance engineer.

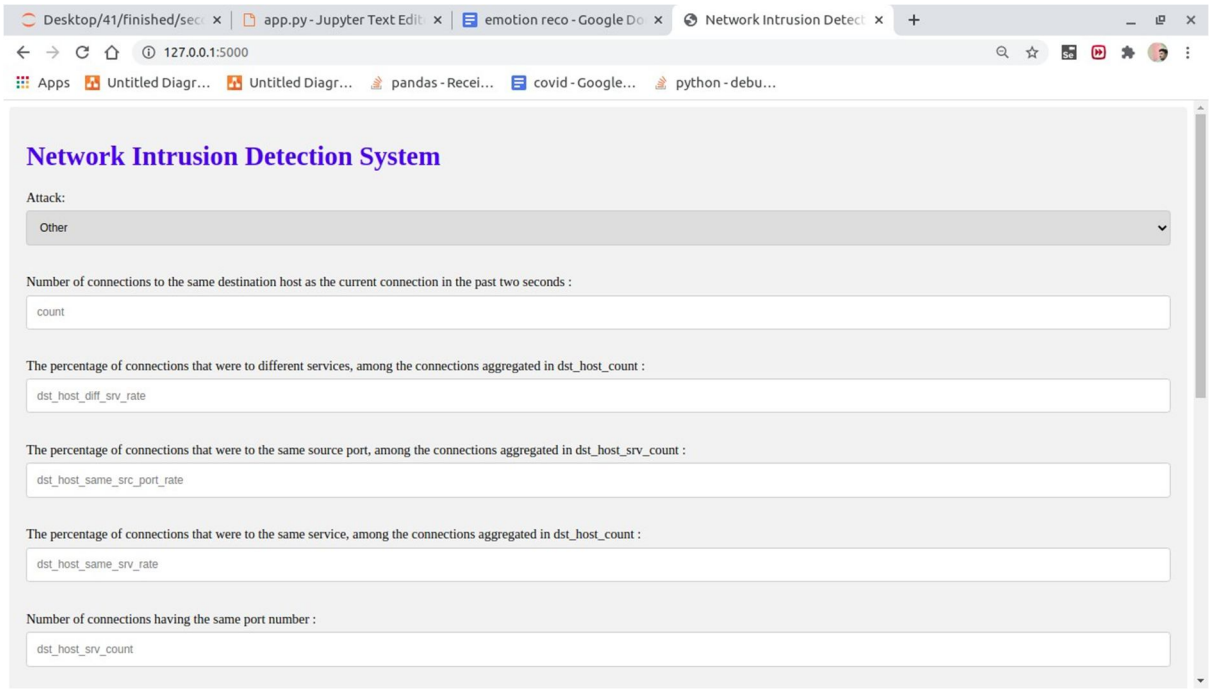
## VII. RESULTS




## 8. Conclusion

We presented a vision-based method for interpreting single hand motions from Sign Language in this paper. To classify the spatial and temporal features, this study used a prediction approach. The geographical features were classified using CNN, whereas the temporal features were classified using RNN. This demonstrates how CNN and RNN may be used to learn

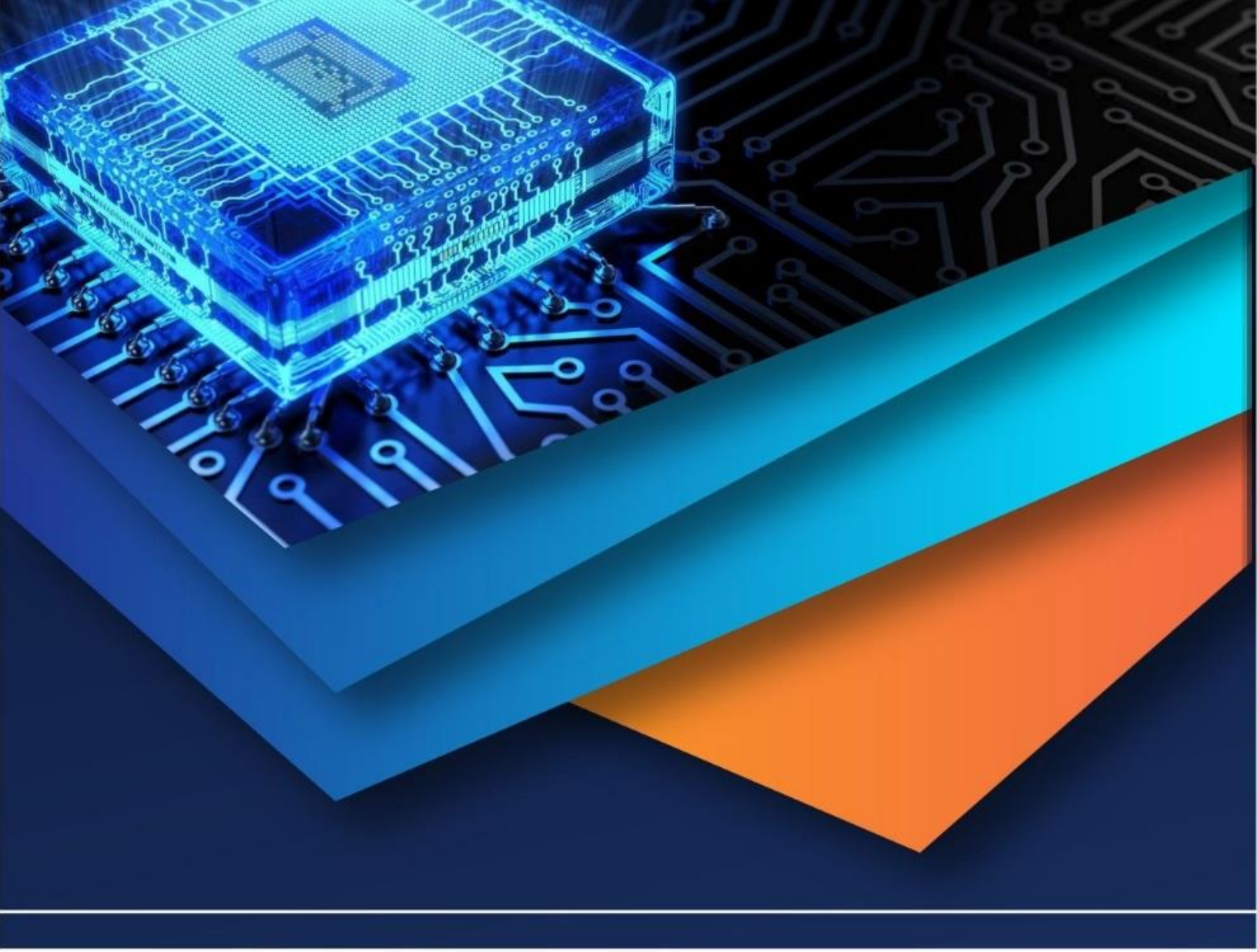



## VIII. ACKNOWLEDGMENT

We express our sincere gratitude to our guide, Assistant Professor Mr. G. Sekhar Reddy for suggestion and support during every stage of this work. We also convey our deep sense of gratitude to Professor Dr. K. S. Reddy, Head of Information Technology department.

## REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.
[2] R. Christopher, “Port scanning techniques and the defense against them,” SANS Institute, 2001.
[3] M. Baykara, R. Das¸, and I. Karado ˘gan, “Bilgi g ¨uvenli ˘gisistemlerindekullanilanarac¸larinincelenmesi,” in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.
[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, “Practical automated detection of stealthy portscans,” Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.
[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, “Surveillance detection in high bandwidth environments,” in DARPA Information Survivability Conference and Exposition, 2003.Proceedings, vol. 1. IEEE, 2003, pp. 130–138.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY