iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VIII **Month of publication:** August 2022

DOI: https://doi.org/10.22214/ijraset.2022.45132

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 VIII Aug 2022- Available at www.ijraset.com

An Efficient Spam Detection Technique for IoT Devices Using Machine Learning

Mrs. S. Mounasri[1], D. Tejaswani[2], A. Mounika[3], S. Bhuvaneshwari[4]
[1]*Assistant Professor, Department of Computer Science Engineering, Sridevi Women's Engineering College, Hyderabad, Telangana*
[2, 3, 4]*Under Graduate Student, Department of Computer Science Engineering Sridevi Women's Engineering College, Hyderabad*

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. The volume of data released from these devices will increase many-fold in the years to come. 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 project, we propose the security of the IoT devices by detecting spam using machine learning. In this framework, five machine learning models are evaluated using various metrics with a large collection of inputs features sets. Each model computes a spam score by considering the refined input features. This score depicts the trustworthiness of IoT device under various parameters. The results obtained proves the effectiveness of the proposed scheme in comparison to the other existing schemes.***

I. INTRODUCTION

A. *Purpose*

The main purpose of this project is to present a thorough and complete assessment of current research on detecting review spam using various machine learning approaches, as well as to develop methodology for further exploration.

Internet of Things (IoT) enables convergence and implementations between the real-world objects irrespective of their geographical locations. IoT applications need to protect data privacy to fix security issues such as intrusions, spoofing attacks, DoS attacks, DoS attacks, jamming, eavesdropping, spam, and malware.

B. *Scope*

The main purpose of this project is to present a thorough and complete assessment of current research on detecting review spam using various machine learning approaches, as well as to develop methodology for further exploration. Internet of Things (IoT) enables convergence and implementations between the real-world objects irrespective of their geographical locations. Implementation of such network management and control make privacy and protection strategies utmost important and challenging in such an environment. IoT applications need to protect data privacy to fix security issues such as intrusions, spoofing attacks, DoS attacks, DoS attacks, jamming, eavesdropping, spam, and malware. For example, wearable devices collect and send user's health data to a connected smartphone should prevent leakage of information 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

C. *Model Diagram/Overview*

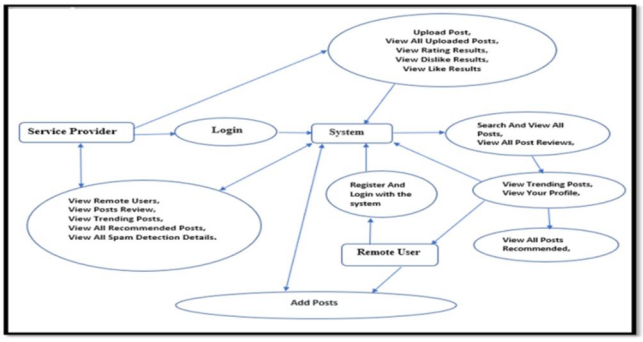


Fig. Model Diagram

The above model diagram depicts the information how the remote user and server is connected and how can we see the reviews of the posts.

II. SYSTEM ANALYSIS

A. *Existing System*

1) *Denial of Service (DDoS) Attacks:* The attackers can flood the target database with unwanted requests to stop IoT devices from having access to various services. DDoS can exhaust all the resources provided by the service provider.
2) *RFID Attacks:* These are the attacks imposed at the physical layer of IoT device. The common attacks possible at the sensor node are attacks on availability, attacks on authenticity, attacks on confidentiality, Cryptography keys brute-forcing.
3) *Internet Attacks:* The IoT device can stay connected with Internet to access various resources. The spammers who want to steal other systems information or want their target website to be visited continuously, use spamming techniques.
4) NFC Attacks: These attacks are mainly concerned with electronic payment frauds. The possible attacks are unencrypted traffic, Eavesdropping, and Tag modification.
5) The solution for this problem is the conditional privacy protection. So, the attacker fails to create the same profile with the help of user's public key. This model is based on random public keys by trusted service manager.

➢ *Disadvantages Of Existing System*

- In the existing work, the system is less effective due to lack of Spam Detection in IoT using Machine Learning framework.
- This system is less performance in which it is clear that Supervised machine learning techniques is absence.

B. *Problem Statement*

1) IoT applications need to protect data privacy to fix security issues such as intrusions, spoofing attacks, DoS attacks, DoS attacks, jamming, eavesdropping, spam, and malware.
2) The safety measures of IoT devices depends upon the size and type of organization in which it is imposed.

C. *Proposed System*

1) The digital world is completely dependent upon the smart devices. The information retrieved from these devices should be spam free.
2) The information retrieval from various IoT devices is a big challenge because it is collected from various domains. As there are multiple devices involved in IoT, so a large volume of data is generated having heterogeneity and variety.
3) Here support vector machine is used to detect the spam in particular IoT devices.
4) We can call this data as IoT data. IoT data has various features such as real-time, multi-source, rich and sparse.
5) Here we use Randomforest algorithm to depicts the trustworthiness of IoT device under various parameters.

- The proposed scheme of spam detection is validated using five different machine learning models.
- An algorithm is proposed to compute the spamicity score of each model which is then used for detection and intelligent decision making.
- Based upon the spamicity score computed in previous step, the reliability of IoT devices is analyzed using different evaluation metrics.

The target is to resolve the issues in the IoT devices deployed within home. But, the proposed methodology considers all the parameters of data engineering before validating it with machine learning models.

➢ *Advantages Of Proposed System*

- The proposed scheme of spam detection is validated using five different machine learning models.
- An algorithm is proposed to compute the spam city score of each model which is then used for detection and intelligent decision making.
- Based upon the spam city score computed in previous step, the reliability of IoT devices is analyzed using different evaluation metrics.

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

III. SYSTEM REQUIREMENT SPECIFICATION

A. *Functional Requirements*

1) Maintenance: The system should associate a supervisor indicator with each job class.
2) Changing Dues in the System: The system should handle any number of fees (existing and new) associated with unions.
3) The system should capture and maintain job class status (i.e., active or inactive). Some job classes are old and are no longer used. However, they still need to be maintained for legal, contract and historical purposes.
4) The system should assign the Supervisor Code based on the value in the Job Class table and additional criteria as specified by the clients.
5) The system should provide the Labor Relations office with the ability to override the system-derived Bargaining Unit code and the Union Code for to-be-determined employee types, including hourly appointments.

B. *Non Functional Requirements*

1) *Usability:* Prioritize the important functions of the system based on usage patterns. Frequently used functions should be tested for usability, as should complex and critical functions.
2) *Reliability:* Reliability defines the trust in the system that is developed after using it for a period of time. It defines the likeability of the software to work without failure for a given time period.
3) *Performance:* What should system response times be, as measured from any point, under what circumstances? Are there specific peak times when the load on the system will be unusually high?
4) *Supportability:* The system needs to be cost-effective to maintain. Maintainability requirements may cover diverse levels of documentation, such as system documentation, as well as test documentation.

C. *Hardware Requirements*

Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user.
Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

1) Processor: Pentium –IV
2) RAM: 4 GB (min)
3) Hard Disk: 20 GB
4) Key Board: Standard Windows Keyboard
5) Mouse: Two or Three Button Mouse

D. *Software Requirements*

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.
The appropriation of requirements and implementation constraints gives the general overview of the project in regards to what the areas of strength and deficit are and how to tackle them.

1) *Operating system:* Windows 7 Ultimate.
2) *Coding Language:* Python.
3) *Front-End:* Python.
4) *Back-End:* Django-ORM
5) *Designing:* Html, CSS, Javascript
6) *Database:* MySQL(WAMP Server)IV. SYSTEM DESIGN

IV. SYSTEM DESIGN

A. *System Architecture*

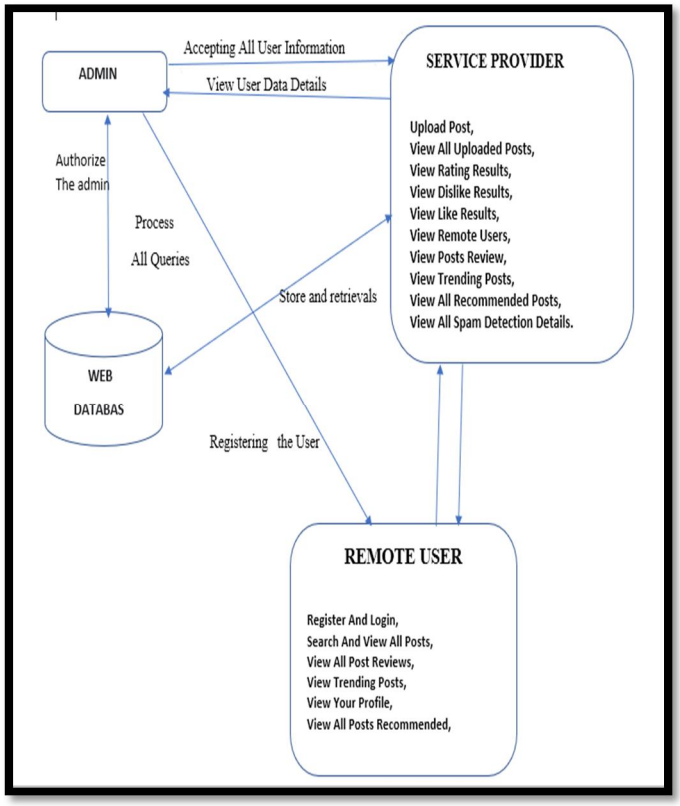


Fig System Architecture

System architecture refers to the placement of these software components on physical machines. Two closely related components can be co-located or placed on different machines. The location of components will also impact performance and reliability. The resulting architectural style ultimately determines how components are connected, data is exchanged, and how they all work together as a coherent system.

1) *Machine Learning Models*

Model no.	Model	Method	Package	Tuning parame-ters
Model1	Bagged Model	Bag	Caret	Vars
Model2	Bayesian Generalized Linear Model	bayesglm	Arm	None
Model3	Boosted Linear Model	BstLm	bst, plyr	mstop, nu
Model4	eXtreme Gradient Boosting	xg-bLin-ear	Xgboost	nrounds, lambda, alpha
Model5	Generalized Lin-ear Model with Stepwise Feature Selection	glm-StepAIC	MASS	None

2) Supervised Learning

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans. You can use supervised learning when the output data is known. The algorithm will predict new data.

3) Unsupervised Learning

In unsupervised learning, an algorithm explores input data without being given an explicit output variable (e.g., explores customer demographic data to identify patterns)

You can use it when you do not know how to classify the data, and you want the algorithm to find patterns and classify the data for you.

B. System Components (Modules)

Modules used in resume categorization are:

1) *Service Provide:* In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Upload Post View All Uploaded Posts, View Rating Results, View Dislike Results, View Like Results, View Remote Users, View Posts Review, View Trending Posts, View All Recommended Posts.
2) *View and Authorize Users:* In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.
3) *Remote User:* In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like search and view all posts, view all post reviews, view trending posts, view your profile, view all posts recommended. Extension Offload task time required to execute and plots a bar graph.

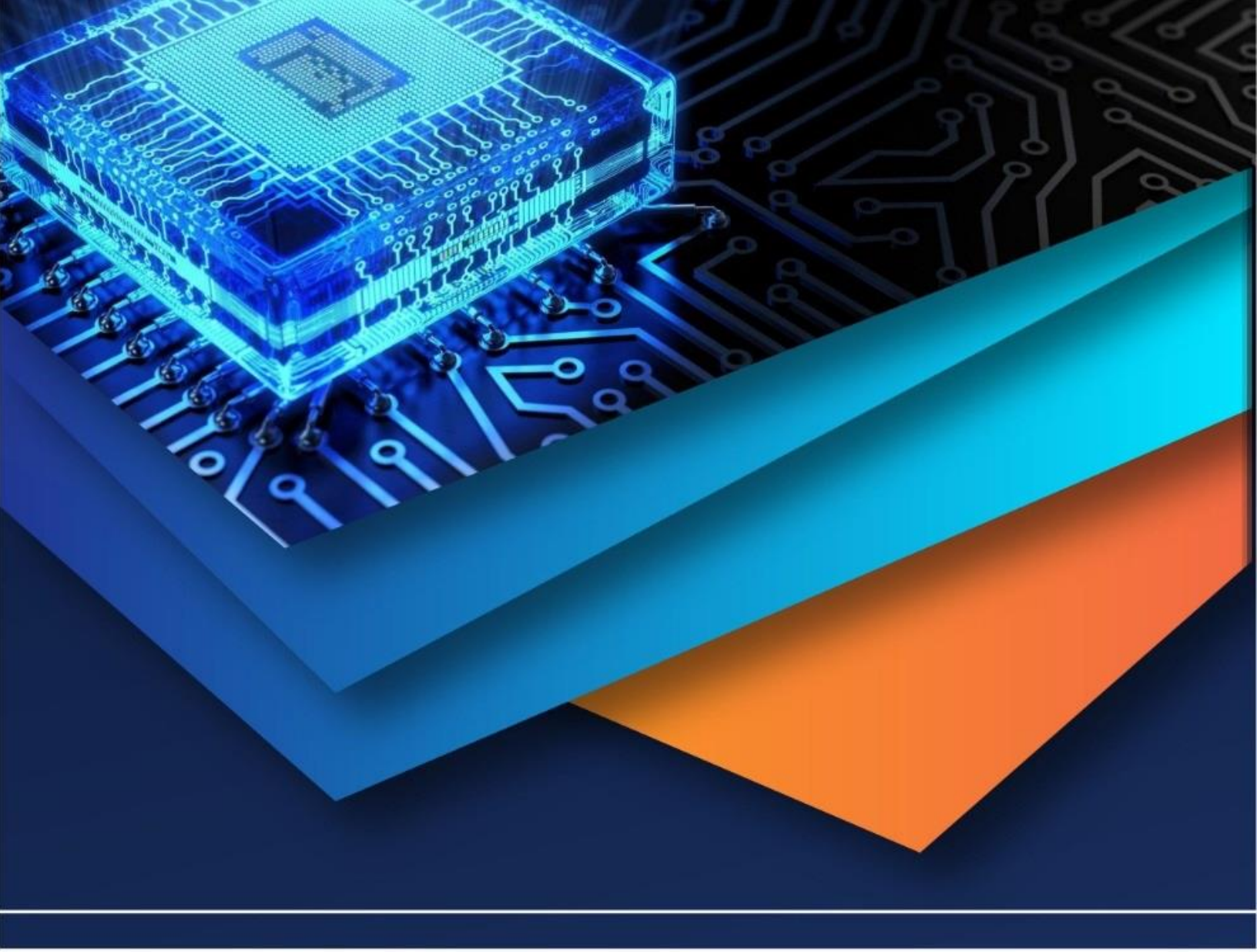
V. CONCLUSION:

The following conclusion can be presented:

The proposed framework, detects the spam parameters of IoT devices using machine learning models. The IoT dataset used for experiments, is pre-processed by using feature engineering procedure. By experimenting the framework with machine learning models, each IoT appliance is awarded with a spam score. The spamicity score is used in this research to determine the reliability of IoT devices in the smart home organisation. Different ML models were utilised to assess the time-arrangement information produced by keen metres through extensive tests and analysis. This refines the conditions to be taken for successful working of IoT devices in a smart home. In future, we are planning to consider the climatic and surrounding features of IoT device to make them more secure and trustworthy.

REFERENCES

[1] Z.-K. Zhang, M. C. Y. Cho, C.-W. Wang, C.-W. Hsu, C.-K. Chen, and S. Shieh, "Iot security: ongoing challenges and research opportunities," in 2014 IEEE 7th international conference on service-oriented computing and applications. IEEE, 2014, pp. 230–234.
[2] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for iot security and privacy: The case study of a smart home," in 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops). IEEE, 2017, pp. 618–623.
[3] E. Bertino and N. Islam, "Botnets and internet of things security," Computer, no. 2, pp. 76–79, 2017.
[4] C. Zhang and R. Green, "Communication security in internet of thing: preventive measure and avoid ddos attack over iot network," in Proceedings of the 18th Symposium on Communications & Networking. Society for Computer Simulation International, 2015, pp. 8–15.
[5] W. Kim, O.-R. Jeong, C. Kim, and J. So, "The dark side of the internet: Attacks, costs and responses," Information systems, vol. 36, no. 3, pp. 675–705, 2011.
[6] H. Eun, H. Lee, and H. Oh, "Conditional privacy preserving security protocol for nfc applications," IEEE Transactions on Consumer Electronics, vol. 59, no. 1, pp. 153–160, 2013
[7] R. V. Kulkarni and G. K. Venayagamoorthy, "Neural network based secure media access control protocol for wireless sensor networks," in 2009 International Joint Conference on Neural Networks. IEEE, 2009, pp. 1680–1687.
[8] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 1996– 2018, 2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY