IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 8, August 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 8, August 2024

|DOI: 10.15680/IJIRSET.2024.1308086|

An Effective and Efficient Deep Learning Methodology for Detecting of Spam in IOT Network using RNN LSTM Technique

MD Imteyaz Ahmad[1], Prof. Sarwesh Site[2]

M. Tech. Scholar, Department of Computer Science and Engineering, All Saint's College of Technology, Bhopal, India[1]

Assistant Professor, Department of Computer Science and Engineering, All Saint's College of Technology, Bhopal, India[2]

ABSTRACT: This paper presents an efficient deep learning technique for detecting of spam in the IOT network. The RNN-LSTM techniques is applied and optimized for better performance than others. For each topic, the existing problems are analyzed, and then, current solutions to these problems are presented and discussed. The simulation results show that the proposed sentiment analysis method has higher precision, recall and F1 score. The method is proved to be effective with high accuracy. The simulation and analysis are done using the python spyder 3.7 software. The overall accuracy achieved by the proposed work is 95.72 % while previous it is achieved 91.8 %. The error rate of proposed technique is 4.28 % while 8.2 % in existing work. Therefore, it is clear from the simulation results; the proposed work is achieved significant better results than existing work.

KEYWORDS: Spam, Deep Learning, Recurrent Neural Network (RNN), Long Short Term Memory (LSTM)

I. INTRODUCTION

Internet of Things (IoT) enables convergence and implementations between the real-world objects irrespective of their geographical locations. Implementation of IOT network management and control makes 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.

The safety measures of IoT devices depend upon the size and type of organization in which it is imposed. The behaviour of users forces the security gateways to cooperate. In other words, we can say that the location, nature, application of IoT devices decides the security measures. For instance, the smart IoT security cameras in the smart organization can capture the different parameters for analysis and intelligent decision making. The maximum care to be taken is with web based devices as maximum number of IoT devices are web dependent. It is common at the workplace that the IoT devices installed in an organization can be used to implement security and privacy features efficiently. 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. The expanding nature of IoT attracts both the audience, i.e., the users and the attackers.

However, with the emergence of ML in various attacks scenarios, IoT devices choose a defensive strategy and decide the key parameters in the security protocols for trade-off between security, privacy and computation. This job is challenging as it is usually difficult for an IoT system with limited resources to estimate the current network and timely attack status.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 8, August 2024

|DOI: 10.15680/IJIRSET.2024.1308086|

II. SPAM

The Spamming is the use of messaging systems to send multiple unsolicited messages (spam) to large numbers of recipients for the purpose of commercial advertising, for the purpose of non-commercial proselytizing, for any prohibited purpose (especially the fraudulent purpose of phishing), or simply repeatedly sending the same message to the same user. While the most widely recognized form of spam is email spam, the term is applied to similar abuses in other media: instant messaging spam, Usenet newsgroup spam, Web search engine spam, spam in blogs, wiki spam, online classified ads spam, mobile phone messaging spam, Internet forum spam, junk fax transmissions, social spam, spam mobile apps [1]. television advertising and file sharing spam. It is named after Spam, a luncheon meat, by way of a Monty Python sketch about a restaurant that has Spam in almost every dish in which vikings annoyingly sing "Spam" repeatedly [2].

Spamming remains economically viable because advertisers have no operating costs beyond the management of their mailing lists, servers, infrastructures, IP ranges, and domain names, and it is difficult to hold senders accountable for their mass mailings. The costs, such as lost productivity and fraud, are borne by the public and by Internet service providers, which have added extra capacity to cope with the volume. Spamming has been the subject of legislation in many jurisdictions [3].

Email Spam- Email spam, also known as unsolicited bulk email (UBE), or junk mail, is the practice of sending unwanted email messages, frequently with commercial content, in large quantities. Spam in email started to become a problem when the Internet was opened for commercial use in the mid-1990s. It grew exponentially over the following years, and by 2007 it constituted about 80% to 85% of all e-mail, by a conservative estimate. Pressure to make email spam illegal has resulted in legislation in some jurisdictions, but less so in others.

Messaging spam- Instant messaging spam makes use of instant messaging systems. Although less prevalent than its e-mail counterpart, according to a report from Ferris Research, 500 million spam IMs were sent in 2003, twice the level of 2002.

Newsgroup Spam and **Forum spam-** Newsgroup spam is a type of spam where the targets are Usenet newsgroups. Spamming of Usenet newsgroups actually pre-dates e-mail spam. Usenet convention defines spamming as excessive multiple posting, that is, the repeated posting of a message (or substantially similar messages). The prevalence of Usenet spam led to the development of the Breitbart Index as an objective measure of a message's "spamminess".

Mobile Phone Spam- Mobile phone spam is directed at the text messaging service of a mobile phone. This can be especially irritating to customers not only for the inconvenience, but also because of the fee they may be charged per text message received in some markets. To comply with CAN-SPAM regulations in the US, SMS messages now must provide options of HELP and STOP, the latter to end communication with the advertiser via SMS altogether.

Despite the high number of phone users, there has not been so much phone spam, because there is a charge for sending SMS. Recently, there are also observations of mobile phone spam delivered via browser push notifications. These can be a result of allowing websites which are malicious or delivering malicious ads to send user notifications.

Social Networking Spam- Facebook and Twitter are not immune to messages containing spam links. Spammers hack into accounts and send false links under the guise of a user's trusted contacts such as friends and family. As for Twitter, spammers gain credibility by following verified accounts such as that of Lady Gaga; when that account owner follows the spammer back, it legitimizes the spammer. Twitter has studied what interest structures allow their users to receive interesting tweets and avoid spam, despite the site using the broadcast model, in which all tweets from a user are broadcast to all followers of the user.

Spam in blogs- Blog spam is spamming on weblogs. In 2003, this type of spam took advantage of the open nature of comments in the blogging software Movable Type by repeatedly placing comments to various blog posts that provided nothing more than a link to the spammer's commercial web site. Similar attacks are often performed against wikis and guestbooks, both of which accept user contributions. Another possible form of spam in blogs is the spamming of a certain tag on websites such as Tumblr.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 8, August 2024

|DOI: 10.15680/IJIRSET.2024.1308086|

VoIP spam- VoIP spam is VoIP (Voice over Internet Protocol) spam, usually using SIP (Session Initiation Protocol). This is nearly identical to telemarketing calls over traditional phone lines. When the user chooses to receive the spam call, a pre-recorded spam message or advertisement is usually played back. This is generally easier for the spammer as VoIP services are cheap and easy to anonymize over the Internet, and there are many options for sending mass number of calls from a single location. Accounts or IP addresses being used for VoIP spam can usually be identified by a large number of outgoing calls, low call completion and short call length.

Mobile Apps Spam- Spamming in mobile app stores include (i) apps that were automatically generated and as a result do not have any specific functionality or a meaningful description; (ii) multiple instances of the same app being published to obtain increased visibility in the app market; and (iii) apps that make excessive use of unrelated keywords to attract users through unintended searcher.

III. PROPSOED METHODOLOGY

The main contributions of this work will be summarized as follows.

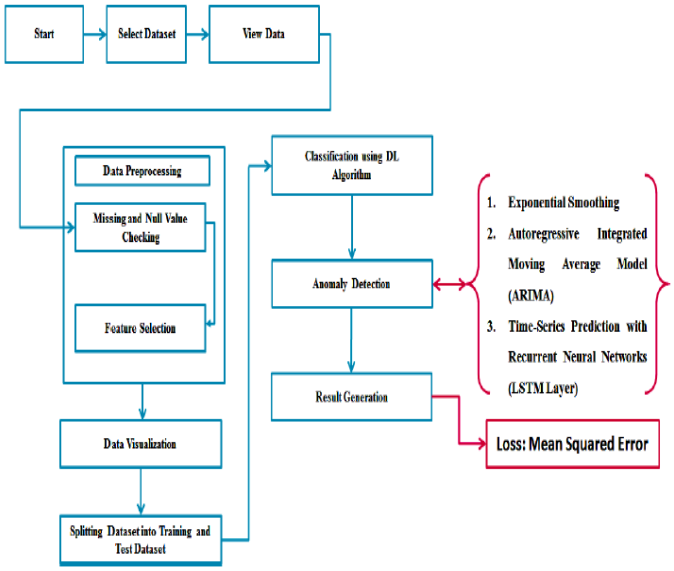

Figure 1: Flow Chart

- To collect SPAM dataset from kaggle website.
- To implement proposed approach based on machine learning technique.
- To improve the performance of Spam Detection in IoT using proposed technique.
- To simulate proposed method on spyder python 3.7 software.
- To enhance the performance of overall prediction result.

Steps-
1. Firstly, download the SPAM dataset from kaggle website, which is a large dataset provider and machine learning repository Provider Company for research.
2. Now apply the preprocessing of the data, here handing the missing data, removal null values.
3. Now extract the data features and evaluate in dependent and independent variable.
4. Now apply the classification method based on the deep learning recurrent neural network and long term short memory (RNN-LSTM).
5. Now generate confusion matrix and show all predicted class like true positive, false positive, true negative and false negative.
6. Now calculate the performance parameters by using the standard formulas in terms of the precision, recall, F_measure, accuracy and error rate.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 8, August 2024

|DOI: 10.15680/IJIRSET.2024.1308086|

The proposed model shows the main steps for preprocessing stage, feature extraction, and classification. The main objective of this process is to detect spam in IoT devices using machine learning technique. For that the first process is to pre-process the dataset to remove missing values and null values from the taken smart home dataset. In order to classify spams, we need to record data from IoT devices and then process them to extract different features.
The data sets are made from the features and then we classify the dataset. In this process we propose deep learning (RNN-LSTM) algorithms to classify the data. Finally it improves the accuracy of detection spam from IoT Devices.

Data Selection and Loading
The data selection is the process of selecting the data for predicting the depression patient emotion from the IoT smart home dataset. It contains the readings with a time span of 1 minute of house appliances in kW from a smart meter and weather conditions of that particular region.

Data Preprocessing
Data pre-processing is the process of removing the unwanted data from the dataset.
- Missing data removal
- Encoding Categorical data

Missing data removal: In this process, the null values such as missing values are removed using imputer library.

Data Visualization
Exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

Splitting Dataset Into Train And Test Data
Data splitting is the act of partitioning available data into two portions, usually for cross validator purposes. One portion of the data is used to develop a predictive model and the other to evaluate the model's performance. Separating data into training and testing sets is an important part of evaluating data mining models.

IV. SIMUALTION RESULT

Simulation Tool: **Python-** It is an interpreter, raised level, comprehensively helpful programming language.
Made by Guido van Rossum and first delivered in 1991, Python's plan reasoning stresses code clarity with its famous utilization of critical whitespace. Its language builds and article organized philosophy hope to help developers with forming clear, genuine code for litt le and colossal scale projects.

Results: Figure 2 is presenting various IOT devices total count with the month format. The month consider from the jan to dec.

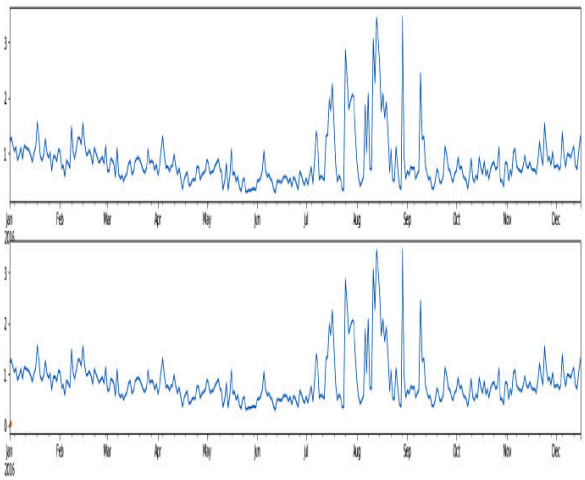

Figure 2: Training

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 8, August 2024

|DOI: 10.15680/IJIRSET.2024.1308086|

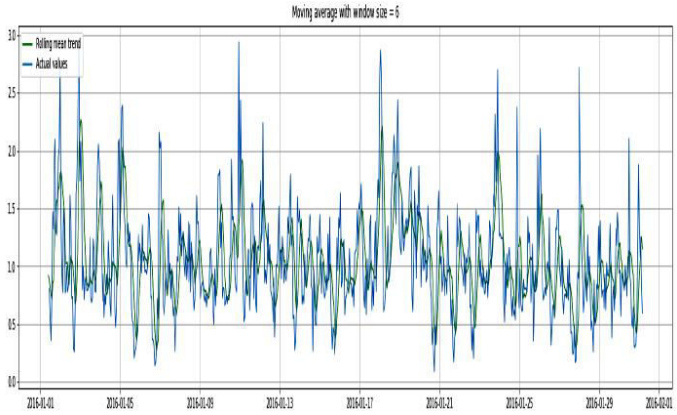

Figure 3: Moving size 6

Figure is 3 is showing actual and the average value during the moving average window size 6.
Figure is 4 is showing actual and the average value during the moving average window size 24. The upper and lower bond range is also mentioned.

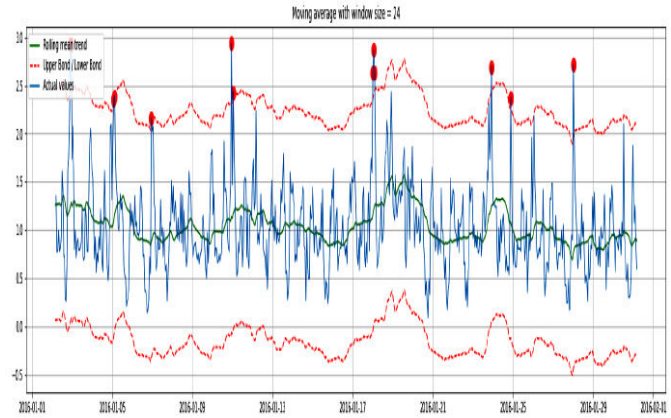

Figure 4: Moving size 24

Figure is 5 is showing predicted and actual data from the given dataset. It is clear from the result graph; the most of the data is predicted.

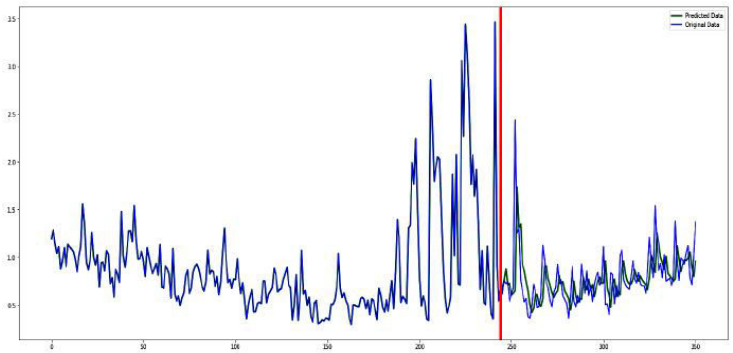

Figure 5: Predicted and actual data

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 8, August 2024

|DOI: 10.15680/IJIRSET.2024.1308086|

Table 1: Simulation Results

Sr. No.	Parameter Name	Value
1	Method	Deep Learning (RNN-LSTM)
2	Accuracy	95.72 %
3	Classification error	4.28 %
4	MSE	0.076%
5	Memory	100 MB

V. CONCLUSION

The proposed framework, detects the spam parameters of IoT devices using machine deep learning models. The IoT dataset used for simulation is pre-processed by using feature engineering procedure. By simulation the framework with deep learning models, each IoT appliance is awarded with a spam score. This refines the conditions to be taken for successful working of IoT devices in a smart home. This dissertation presents an efficient spam detection technique for IOT devices using deep learning approach. The simulation is performed using Python spyder environment, simulated results shows that the overall accuracy achieved by the proposed work is 95.72 % while previous it is achieved 91.8 %. The error rate of proposed technique is 4.28 % while 8.2 % in existing work. Therefore it is clear from the simulation results; the proposed work is achieved significant better results than existing work.

REFERENCES

1. Makkar, S. Garg, N. Kumar, M. S. Hossain, A. Ghoneim and M. Alrashoud, "An Efficient Spam Detection Technique for IoT Devices Using Machine Learning," in IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp. 903-912, Feb. 2021, doi: 10.1109/TII.2020.2968927.
2. F. Hossain, M. N. Uddin and R. K. Halder, "Analysis of Optimized Machine Learning and Deep Learning Techniques for Spam Detection," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1-7, doi: 10.1109/IEMTRONICS52119.2021.9422508.
3. A. Makkar, U. Ghosh, P. K. Sharma and A. Javed, "A Fuzzy-based approach to Enhance Cyber Defence Security for Next-generation IoT," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3053326.
4. G. Fortino, F. Messina, D. Rosaci and G. M. L. Sarne, "ResIoT: An IoT social framework resilient to malicious activities," in IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 5, pp. 1263-1278, September 2020, doi: 10.1109/JAS.2020.1003330.
5. K. A. Al-Thelaya, T. S. Al-Nethary and E. Y. Ramadan, "Social Networks Spam Detection Using Graph-Based Features Analysis and Sequence of Interactions Between Users," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 206-211, doi: 10.1109/ICIoT48696.2020.9089509.
6. T. Y. Ho, W. Chen, M. Sun and C. Huang, "Visualizing the Malicious of Your Network Traffic by Explained Deep Learning," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 687- 692, doi: 10.1109/ICAIIC48513.2020.9065247.
7. J. Zhang, Q. Yan and M. Wang, "Evasion Attacks Based on Wasserstein Generative Adversarial Network," 2019 Computing, Communications and IoT Applications (ComComAp), 2019, pp. 454-459, doi: 10.1109/ComComAp46287.2019.9018647.
8. K. Singh, S. Bhushan and S. Vij, "Filtering spam messages and mails using fuzzy C means algorithm," 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), 2019, pp. 1-5, doi: 10.1109/IoTSIU. 2019.8777483.
9. T. Lange and H. Kettani, "On Security Threats of Botnets to Cyber Systems," 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), 2019, pp. 176-183, doi: 10.1109/SPIN.2019.8711780.
10. T. Qiu, H. Wang, K. Li, H. Ning, A. K. Sangaiah and B. Chen, "SIGMM: A Novel Machine Learning Algorithm for Spammer Identification in Industrial Mobile Cloud Computing," in IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2349- 2359, April 2019, doi: 10.1109/TII.2018.2799907.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 8, August 2024

|DOI: 10.15680/IJIRSET.2024.1308086|

11. G. Kumar and V. Rishiwal, "Statistical Analysis of Tweeter Data Using Language Model With KLD," 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU), 2018, pp. 1-6, doi: 10.1109/IoTSIU. 2018.8519938.
12. E. Anthi, L. Williams and P. Burnap, "Pulse: An adaptive intrusion detection for the Internet of Things," Living in the Internet of Things: Cybersecurity of the IoT - 2018, 2018, pp. 1-4, doi: 10.1049/cp.2018.0035.
13. A. Kaushik and S. Talati, "Securing IoT using layer characterstics," 2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2017, pp. 290-298, doi: 10.1109/ICATCCT.2017.8389150.
14. İ. Ü. Oğul, C. Özcan and Ö. Hakdağlı, "Fast text classification with Naive Bayes method on Apache Spark," 2017 25th Signal Processing and Communications Applications Conference (SIU), 2017, pp. 1-4, doi: 10.1109/SIU.2017.7960721.
15. Z. Lv, J. Lloret, H. Song, J. Shen and W. Mazurczyk, "Guest Editorial: Secure Communications Over the Internet of Artificially Intelligent Things," in IEEE Internet of Things Magazine, vol. 5, no. 1, pp. 58-60, March 2022, doi: 10.1109/MIOT.2022.9773087.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details