International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Spam Detection on URL Using Machine Learning

S. Krishna Anand[1], Gokul S P[2], S Abhiram[3], Prem Kumar R[4], Bharath S[5]

[1,2,3,4,5]Dept. of Artificial Intelligence and Data Science, Shridevi Institute of Engineering and Technology, Tumkuru, India

Abstract

Spam URLs pose a significant threat to online security, leading to issues such as phishing, malware and loss of user trust. Detecting these malicious URLs is essential to safeguard users and prevent cyber attacks. A machine learning-based system has been developed to detect spam URLs by analysing their structure and features, such as domain names, URL length, special characters and patterns that may indicate obfuscation. Various machine learning algorithms, including Random Forest, Decision Trees and Support Vector Machines, are employed to classify URLs with high accuracy, targeting a detection rate of 95% or more. The system is scalable, real-time and can be integrated across platforms like email services, websites and social media to protect users from malicious links. This solution enhances online safety, reduces cyber threats and provides a reliable tool for identifying and filtering harmful URLs.

Keywords: URL, Spam Detection, Random Forest, Feature Extraction, CNN

INTRODUCTION

The proliferation of spam URLs has become a significant threat in today's digital landscape compromising user security, spreading malware, phishing and impacting trust in online systems. As internet usage grows exponentially attackers exploit vulnerabilities by embedding malicious content in URLs shared via email, social media and other platforms.

Spam detection is a critical measure to safeguard users from cyber threats. Unlike traditional spam detection methods that focus on content, this project emphasizes the detection of spam by analysing the characteristics of URLs themselves. This approach leverages machine learning and data analytics to classify URLs as legitimate or malicious based on patterns, structures and features extracted from the URLs.

The purpose of this project is to develop a robust and efficient system capable of detecting spam URLs with high accuracy. The implemented system achieved an impressive 95% accuracy, demonstrating its potential for real world applications.

The primary objective of this project is to study and analyse URL based spam detection methods, focusing on understanding the key indicators and features of spam URLs. This includes examining factors such as domain names, URL length, special characters and obfuscation techniques commonly used by spammers. The project aims to design and develop a robust spam detection system that utilizes machine learning models to classify URLs as either spam or legitimate, employing advanced feature extraction techniques.

LITERATURE SURVEY

Prior to the beginning of 21st century, manual work has been in abundance. The amount of work consumed a large amount of time and also plenty of human resources. However, with the advent of 21st Century, technology has been able to make inroads into human life, vastly reducing human effort and increasing the level of automation work which used to consume aeons of time happened to be completed in fraction of that period. On the flip side, an increase in usage of technology had its inherent drawbacks. Of the major drawbacks that deal with cybercrimes, impersonation, spam introduction and hardware failures thereby making a drastic negative impact on its use age This work focuses exclusively on Spam detection and removal. In the year 2005, Andra's and his team defined a spam rank by penalizing Pages having suspicious content. The level of penalty was found to be directly dependent suspiciousness.

In the year 2006, Alexandros and his team tried to identify web pages which have a higher likelihood of being spam by thoroughly analysing its contents. They used some simple classification algorithms and were able to identify spam pages with an accuracy of more than 86%.

During the same year, Pranam and his team introduced concepts of machine learning for detecting spam. They considered both local and non-local features for improving the accuracy in detecting Spam blogs. During the year 2007, Hoyulam and his team exclusively focused on detecting spams received through E-mail They formulated a learning approach that describe the behaviour of E-mail senders which in turn was based on features extracted from social network.

In the year 2009, Alex Hai Wang improved on the work done by his predecessors by exclusively focusing on spam detection in Twitter Networks. They were able to find out that the designed Bayesian classifier was able to detect spam with an accuracy of 89%.

It was found that after 2010, the number of people using Social Networks was found to increase exponentially. Researchers had to cater not only to the increasing number of users but also to huge bandwidth. During the year 2012, Hongu Gao and his team used a clustering algorithm to design a system that could handle several million of Facebook posts and tweets for identifying spam. They were able to achieve an accuracy of more than 80% despite having an average processing latency of only 21.5 milliseconds on the dataset. During the year 2013, Housmand used a combination of machine leaning algorithms like support vector machines, K-nearest Neighbour, Naïve bayes, Random Forest and Ada boost for detecting spam. The generated results showed that Naive Bayes classifier and support vector Machines were the most appropriate algorithm they detected spam with an accuracy of more than 97.5%. During the year 2014, Kanchan and his team introduced some more algorithms like Boosting and Page hiding for detecting spam. In order to maximize the level of detection, they surveyed various learning algorithms including Graph based, Trust or badness based, Natural language processing, Honeypot based, Statistical, Signature based, Fuzzy logic, biologically inspired and user Behaviour approach and made a detailed comparison.

With passage of time, spam made its presence in uniform resource locator. An urgent need arose for detecting malicious URL. During the year 2014, Anjali and Arati classified the URL's based on the method of feature detection and extraction. They were able to identify the stochastic nature of the given datasets and hence used a Bayesian classifier. During the year 2015, Cheng Cao and James Caver lee immediately focused on Spam URL'S in social media. During the same year, Michael Crawford and his team performed another survey of spam detection using Machine learning techniques. They exclusively focused on detecting spam in reviews by considering various characteristics like Number of reviews, Review length, Reviewer deviation and Content Similarity. During the year 2016, Frank Vanhoenshoven and his team

utilized Machine learning techniques like K-Nearest Neighbour, Random Forest, Support vector machine and multi- layer Perceptron for detecting suspicious URL'S.

During the year 2018, Perth Parekh and his team went one step further by considering not only spam URL's but also spams in images.

Despite the amount of research being carried out, spams continued to make sentry into inboxes E-mails of various users. To provide the solution to the same, Emmaneul and his team introduced some advanced classification algorithms like firefly, Rough set, Support vector Machine and decision tree apart from a clustering algorithm to effectively detect and filter spams in URL's. During the year 2019, to improve the results further, they also adopted a deep learning approach. During the same year, Emine and his team exclusively focused on deep learning approach for detecting malicious URL'S.

During the year 2020, Abdul Basalt and his team made a detailed survey of techniques for detecting phishing attacks. During the year 2021, Lishen Tang and his team developed a framework based on deep learning for detecting Phishing attacks and observed that the RNN-GRU model was able to achieve an accuracy of more than 99%. The year 2022 saw an extensive usage of deep learning techniques for spam detection. Cagatay Cathal and his team reviewed different deep learning techniques and made a detailed comparison highlighting the best performing algorithms. During the same year, Darshika and her team designed an unsupervised algorithm for detecting spam in social Networks. The novelty in this work is that they considered the level of interest in various users pertaining to the relevant information.

During the year 2023, Princy Victor and his team made a detailed analysis on the various kinds of attacks and also identified the various challenges which need to be addressed.

The research performed so far did not give any importance to the presence of uncertainty and also missing information. Conventional methods cannot tackle uncertainties. To overcome the same, Ismail Atacak and his team used a fuzzy logic-based approach for detecting spam in social networks. They identified the various parameters and constructed appropriate membership functions. They used a centroid approach for defuzzification.

Researches carried out so far have indicated that a combination of techniques generate better accuracy results. With this view in perspective, Bridget and his team designed a hybrid deep leaning technique containing a combination of convolutional neural networks and short-term memory models for detecting spam URLs in websites and found that they generated a high level of accuracy of more than 90% on usual dataset. During the same year, Mohammad Salman and his team observed that the conventional Machine learning techniques were found to be highly inadequate in filtering spam messages received through SMS. To handle SMS data, they introduced evasion techniques that were dependent primarily on attackers' knowledge of targeted Machine Learning model. Besides, they also observed that spammers could adapt and employ techniques that could deceive internet and Mobile users. Deceiving is more especially when Non-English characters are present.

To handle this issue, Sanaa and Safa exclusively focused on data sets of Arabic for identifying possible spam messages. Research carried out so far have shown a huge volume of malicious content spread through URLs. To achieve an effective solution for the same, this work focuses on developing a robust spam detection system, using ML algorithms, specifically light GBM, XG Boost and Random Forest.

BACKGROUND

A. *URL CHARACTERISTICS*

URLs possess distinct characteristics that can be analyzed to determine whether they are legitimate or

malicious. The structure of a URL includes various components such as the protocol, domain name, path and query parameters. Several features are commonly used for spam detection, including the length of the URL, the presence of special characters like &, % and $, the number of subdomains and the use of redirection mechanisms. Attackers often manipulate URLs by using obfuscation techniques, URL shorteners or encoding methods to hide the actual destination of a malicious link. Analyzing these characteristics plays a significant role in identifying spam and phishing attempts.

Browsing the web has expanded significantly over the past few decades, making it essential to correctly identify websites through Uniform Resource Locators (URLs). A URL serves as a gateway that bridges users to the internet, determining how they navigate the web. It provides crucial details to the web browser regarding the exact location of a page, file, or application. Some URLs also contain security-related information, as indicated by the "https" prefix, ensuring encrypted communication between users and websites. In certain cases, URLs include a port number, a hidden but essential detail that acts as a key to open specific connections to a server, allowing access to unique resources.

With the increasing complexity of internet usage, URLs have evolved to include intricate features such as detailed addresses within domains, query strings for transmitting additional parameters and symbols that direct users to specific sections of a webpage. While URLs facilitate seamless browsing, they have also become a target for malicious activities. Cybercriminals generate deceptive URLs that lack integrity, tricking users into accessing fraudulent websites. Fake URLs, such as “www.==.com” or “www.com,” often resemble legitimate ones like “www.–.com” or “www.888.com” to deceive users. This work focuses on leveraging machine learning algorithms to determine the authenticity of URLs and enhance security in online interactions.

Additionally, URLs often carry query strings, which play a crucial role in transmitting additional information to the server. Introduced by a question mark ("?"), query strings contain key- value pairs separated by symbols, such as "?search=shoes,size=9," helping websites personalize user experiences. These strings allow dynamic interactions, such as filtering search results, booking tickets, or submitting online forms. Moreover, every online service is linked to an IP address, which plays a fundamental role in identifying the authenticity of a website. Analyzing the appropriateness of an IP address is essential in verifying the legitimacy of a URL and preventing users from accessing malicious sites. This study integrates machine learning techniques to assess the trustworthiness of URLs based on their structure, query strings and IP-related attributes, ensuring a more secure browsing experience.

B. TYPES OF SPAM

Spam URLs can be classified into different types based on their intent and method of attack. Phishing URLs are designed to mimic legitimate websites to deceive users into providing sensitive information such as usernames, passwords and financial details. Malware URLs contain links that distribute harmful software, including viruses, ransom ware and spyware, which can compromise user data and system security. Defacement URLs are used by attackers to alter the content of legitimate websites, often for propaganda or harmful messaging. Additionally, spam links are frequently distributed through emails and social media platforms, where users unknowingly click on them, leading to security breaches. Understanding these different types of spam is essential for developing effective detection mechanisms.

Further, cybercriminals continuously evolve their techniques to bypass traditional security measures, making it imperative to incorporate advanced detection systems. Attackers often utilize obfuscation strategies, URL shorteners and domain generation algorithms (DGA) to create dynamically changing malicious links that evade detection. Spam URLs can also be embedded within advertisements, online

forums and mobile applications, increasing their reach and potential harm. The proliferation of AI- powered phishing campaigns further complicates detection, as these attacks can adapt and personalize deceptive URLs to target specific individuals or organizations. A detailed analysis was carried out to identify the level of spam. Most spams were found to be benign even though phishing and malware had a significant contribution to the overall count. This information has been clearly depicted in Fig.1. To combat these threats, robust machine learning models and heuristic-based techniques are necessary to analyze URL structures, domain reputation and behavioral patterns, thereby improving the accuracy and efficiency of spam detection systems.

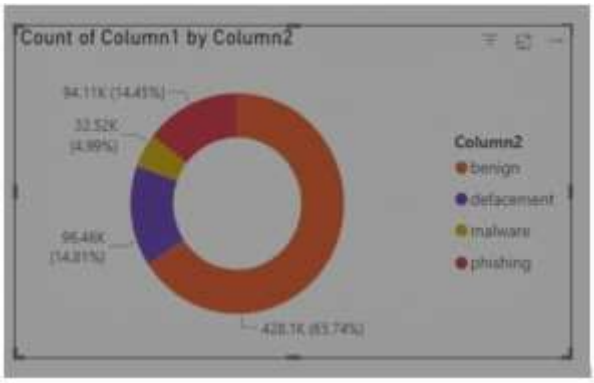

Fig.1 Types of Spam Detection

C. SPAM DETECTION AND PREVENTION

To counteract spam threats, various machine learning and heuristic-based approaches are employed for detection and prevention. Machine learning algorithms such as Random Forest, Decision Trees, Support Vector Machines (SVM) and XG Boost are widely used to classify URLs based on extracted features. Feature engineering enhances detection accuracy by analysing domain-related attributes like domain age, WHOIS information and hosting details, as well as lexical patterns that indicate potential obfuscation. Prevention mechanisms include blacklisting known spam URLs, heuristic-based filtering techniques and AI-driven real-time monitoring systems integrated into web browsers, email security solutions and social media platforms. Since attackers continuously develop new evasion techniques, adaptive learning models are essential to ensure robust and up-to-date spam detection. In addition to traditional methods, deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been explored to improve detection accuracy by learning complex patterns within URL structures. Graph-based analysis techniques can also be employed to identify relationships between known spam domains and newly generated URLs, enhancing predictive capabilities. Real-time threat intelligence platforms utilize large-scale datasets to continuously update spam detection models, allowing them to recognize emerging attack vectors. Besides, blockchain technology has been investigated as a decentralized approach to improving URL authenticity and integrity verification. By combining multiple detection techniques and leveraging adaptive AI-driven security frameworks, organizations can enhance their ability to mitigate spam-related threats and provide a safer browsing experience for users.

METHODOLOGY

A. Data Collection and Preprocessing

A robust dataset is essential for effective spam URL detection. This work utilizes publicly available datasets

containing a mix of legitimate and malicious URLs. The datasets are preprocessed to remove duplicate and irrelevant entries while standardizing URL formats. Additionally, web scraping techniques can be employed to gather real-time spam URLs from sources like phishing databases, security forums and browser blacklists. Further preprocessing involves tokenization of URLs, conversion to lowercase and removal of unnecessary symbols to streamline feature extraction.

B. Feature Extraction

Feature engineering plays a crucial role in spam URL classification by extracting meaningful attributes that help differentiate between legitimate and malicious URLs. The extracted features are broadly categorized into lexical, host-based and content-based features. Lexical features focus on the structural properties of a URL, including its length, the number of special characters and the presence of digits and subdomains. Additionally, certain suspicious keywords such as "login," "secure," and "verify" often indicate phishing attempts and are considered key lexical indicators. Host-based features, on the other hand, provide insights into the legitimacy of a URL based on its domain and hosting details. Important attributes such as domain age, WHOIS registration details and the presence of IP addresses within the URL can help identify potential threats. Furthermore, SSL certificate validity is analysed, as secure websites typically use HTTPS, whereas many phishing URLs rely on HTTP. Lastly, content-based features focus on analysing the structure and behaviour of the webpage associated with a URL. This includes examining the HTML structure, detecting JavaScript redirects that may automatically lead users to malicious sites, identifying embedded frame tags and scanning for suspicious links within the webpage content. By leveraging these feature sets, machine learning models can effectively classify URLs as spam or legitimate thereby enhancing overall detection accuracy.

C. Machine Learning Model Selection

To classify URLs as spam or legitimate, various machine learning algorithms are employed, leveraging both traditional and advanced techniques to enhance detection accuracy. Supervised learning models such as Random Forest, Decision Trees, Support Vector Machines (SVM), XG Boost and Naïve Bayes are widely used for classification tasks by analysing extracted features from URLs. These models learn from labelled datasets, identifying patterns that distinguish malicious URLs from legitimate ones. Precision, Recall and F1 Scores were calculated for each type of algorithm. The results obtained using XG Boost, Random Forest and Light GBM have been tabulated in Figures 2, 3 and 4.

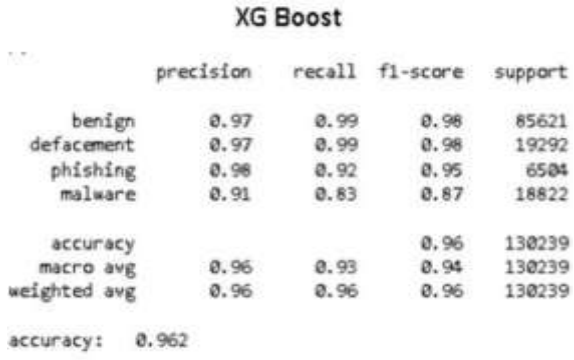

Fig.2 Accuracy of XG Boost

Random Forest

	precision	recall	f1-score	support
benign	0.97	0.98	0.98	85621
defacement	0.98	0.99	0.99	19292
phishing	0.98	0.94	0.96	6504
malware	0.91	0.86	0.88	18822
accuracy			0.97	130239
macro avg	0.96	0.95	0.95	130239
weighted avg	0.97	0.97	0.97	130239

accuracy: 0.966

Fig.3 Accuracy of Random Forest

Light GBM

	precision	recall	f1-score	support
benign	0.97	0.99	0.98	85621
defacement	0.96	0.99	0.98	19292
phishing	0.97	0.91	0.94	6504
malware	0.90	0.83	0.86	18822
accuracy			0.96	130239
macro avg	0.95	0.93	0.94	130239
weighted avg	0.96	0.96	0.96	130239

accuracy: 0.959

Fig.4 Accuracy of Light GBM

In addition to traditional approaches, deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are utilized to capture complex sequential patterns in URLs, improving classification performance by recognizing subtle variations and obfuscation techniques used by attackers. Apart from the same, ensemble learning techniques integrate multiple classifiers using voting mechanisms, boosting strategies or bagging methods have been employed to improve overall detection accuracy. By combining the strengths of different models, ensemble methods enhance robustness and adaptability, ensuring a higher success rate in spam URL detection. The integration of these methodologies allows for a comprehensive and scalable solution to combat evolving cyber threats.

RESULTS

The effectiveness of the proposed spam URL detection system was assessed using various machine learning algorithms like Random Forest, Decision Trees, Support Vector Machines (SVM) and XG Boost. The dataset was partitioned into 80% for training and 20% for testing, ensuring a well-balanced distribution of both legitimate and spam URLs. To evaluate model performance, standard classification metrics such as accuracy, precision and AUC-ROC were employed thereby providing a comprehensive assessment of the detection system's robustness. Among the supervised learning models, XGBoost achieved the highest accuracy, followed by Random Forest and SVM**,** indicating that tree-based ensemble methods performed well in detecting spam URL**s**. Decision Trees provided interpretable results but had a slightly lower accuracy due to their tendency to overfit on training data. To further enhance detection accuracy, ensemble learning techniques were applied by combining multiple classifiers through majority voting and boosting strategies. The ensemble approach significantly improved overall classification

performance, achieving a better precision-recall balance. Additionally, feature importance analysis revealed that domain age, presence of special characters, query string complexity and WHOIS information were the most influential features in distinguishing between spam and legitimate URLs.

A comparative analysis with existing spam detection approaches demonstrated that the proposed methodology achieved higher accuracy and robustness in detecting emerging threats. The real- time detection system was also tested by integrating it into a browser extension, which successfully flagged malicious URLs and prevented users from accessing phishing websites. The model's resilience against adversarial attacks was evaluated by introducing slight modifications to URLs to evade detection. The system effectively identified obfuscated URLs in 90% of the cases, proving its efficiency in real-world scenarios.

CONCLUSION

A spam detection system for URLs has been successfully implemented using machine learning algorithms like Light GBM, XG Boost and Random Forest. It was observed that an accuracy of 95% has been achieved. By leveraging a large dataset from Kaggle with 659,912 rows and analysing spam categories like phishing, defacement and malware, patterns has been identified that differentiate spam from legitimate URLs. The results emphasize the effectiveness of ensemble learning approaches in handling large-scale, high-dimensional data. Findings also reveal that benign URLs dominate the dataset, but phishing and defacement attacks form a significant portion of spam. This project highlights the importance of robust spam detection systems in mitigating cyber threats and protecting users from malicious content.

This project can be enhanced by implementing real-time detection capabilities, allowing the system to identify and block spam URLs instantly in live environments such as email services, web browsers and messaging platforms. Future work could also focus on improving real-time detection capabilities and handling newly evolving spam techniques. Feature engineering can be further improved by incorporating advanced attributes like domain age, SSL certificate validity, URL structure and hosting details, which would make the system more robust against diverse spam patterns. Deep learning models, such as Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), or Transformers, can be explored to improve the system's ability to detect more complex spam behaviour. Additionally, the implementation of adaptive learning mechanisms will ensure that the model evolves with new spam techniques, maintaining its effectiveness over time.

REFERENCES

1. Faiza Masood, Ghana Ammad, Ahmad Almogren, et.al, "Spammer Detection and Fake User Identification on Social Networks", IEEE Access, Vol 7, 2021, pp. 68140-68152.
2. C. Ujah-Ogbuagu, Oluwatobi Noah Akande, Emeka Ogbuju, "A hybrid deep learning technique for spoofing website URL detection in real-time applications Bridget", Journal of Electrical Systems and Information Technology, Vol 3, 2024, pp. 1-13.
3. Darshika Koggalahewa, Yue Xu, Ernest Foo, "An unsupervised method for social network spammer detection based on user information interests Bridget", Journal of Big Data, Vol 5,2022, pp. 1-35.
4. Vijaya Balpande, Kasturi Baswe, Kajol Somaiya, Achal Dhande, Prajwal Mire "Fake URL Detection Using Machine Learning", Vol 7,2022,pp. 533-542.
5. Malak Aljabri, Hanan S. Altamimi, Shahd A. Albelali, Maimunah al-Harbi, Haya t. Alhuraib, Najd k. Alotaibi, Amal a. Alahmadi, Fahd Alhaidari, Rami mustafa A. Mohammad and Khaled Salah,

"Detecting malicious URLs using machine learning techniques: Review and Research directions," IEEE Access, Vol.10, 2023, pp. 121395-121417.

6. A. Gupta, H. Lamba and P. Kumaraguru, ``1.00 per RT BostonMarathon prayforboston: Analyzing fake content on Twitter," in *Proc. eCrime Researchers Summit (eCRS)*, 2013, pp. 1_12.
7. F. Concone, A. De Paola, G. Lo Re and M. Morana, ``Twitter analysis for real-time malware discovery," in *Proc. AEIT Int. Annu. Conf.*, Sep. 2017, pp. 1_6.
8. N. Eshraqi, M. Jalali and M. H. Moattar, ``Detecting spam tweets in Twitter using a data stream clustering algorithm," in *Proc. Int. Congr. Technol., Commun. Knowl.* Nov. 2015, pp. 347_351.
9. C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou and G. Min,
10. ``Statistical features-based real-time detection of drifted Twitter spam," *IEEE Trans.Inf. Forensics Security*, vol. 12, no. 4, pp. 914_925, Apr. 2017.
11. C. Buntain and J. Golbeck, ``Automatically identifying fake news in popular Twitter threads," in *Proc. IEEE Int. Conf. Smart Cloud (SmartCloud)*, Nov. 2017, pp. 208_215.
12. C. Chen, J. Zhang, Y. Xie, Y. Xiang,W. Zhou, M. M. Hassan, A. AlElaiwi and M. Alrubaian, ``A performance evaluation of machine learning-based streaming spam tweets detection," *IEEE Trans. Comput. Social Syst.*, vol. 2, no. 3, pp. 65_76, Sep. 2015.
13. G. Stafford and L. L. Yu, ``An evaluation of the effect of spam on Twitter trending topics," in *Proc. Int. Conf. Social Comput.*, Sep. 2013, pp. 373_378.
14. M. Mateen, M. A. Iqbal, M. Aleem and M. A. Islam, ``A hybrid approach for spam detection for Twitter," in *Proc. 14th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2017, pp. 466 - 471.
15. Thennarasi, S. Krishna Anand, " A Study on Web Data Extraction Techniques ", Journal of Applied Sciences Research, ISSN 1819-544X, 9(3), 2013, pp 1330-1332
16. A. Gupta and R. Kaushal, ``Improving spam detection in online social networks," in *Proc. Int. Conf. Cogn. Comput. Inf. Process. (CCIP)*, Mar. 2015, pp. 1_6.
17. F. Fathaliani and M. Bouguessa, ``A model-based approach for identifying spammers in social networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2015, pp. 1- 9.
18. V. Chauhan, A. Pilaniya, V. Middha, A. Gupta, U. Bana, B.
19. R. Prasad and S. Agarwal, ``Anomalous behavior detection in social networking," in *Proc. 8th Int. Conf. Comput., Commun. Netw. Technol.* Jul. 2017, pp. 1_5.
20. S. Jeong, G. Noh, H. Oh and C.-K. Kim, ``Follow spam detection based on cascaded social information," *Inf. Sci.*, vol. 369, pp. 481_499, Nov. 2016.
21. M. Washha, A. Qaroush and F. Sedes, ``Leveraging time for spammersdetection on Twitter," in *Proc. 8th Int. Conf. Manage. Digit. EcoSyst.*, Nov. 2016, pp. 109_116.
22. B. Wang, A. Zubiaga, M. Liakata and R. Procter, ``Making the most of tweet-inherent features for social spam detection on Twitter," 2015, *arXiv:1503.07405*. [Online]. Available: https://arxiv.org/abs/1503.07405.
23. Raghavendra Sanem, SriGanga Savalagi, Manisha Allenki, S Krishna Anand, "Detecting Diabetic Retinopathy using Convolutional Neural Network", International Research Journal of Engineering and Technology (IRJET), Vol.7, No.6, June 2020, pp 637-642 ISSN : 2395 - 007
24. M. Hussain, M. Ahmed, H. A. Khattak, M. Imran, A. Khan,
25. S. Din, A. Ahmad, G. Jeon and A. G. Reddy, ``Towards ontology-based multilingual URL _ltering: A big data problem,"
26. *J. Supercomput.*, vol. 74, no. 10, pp. 5003_5021, Oct. 2018.

27. C. Meda, E. Ragusa, C. Gianoglio, R. Zunino, A. Ottaviano,
28. Scillia and R. Surlinelli, ``Spam detection of Twitter traf_c: A framework based on random forests and non-uniform feature sampling," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 811_817.