iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36764

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: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Secured Mail Transformation System Using Machine Learning

Rahul Anandpara[1], Dr. Mrs Ranu R. Tuteja[2], Prof. Mrs. Ruchita Kale[3]
[1, 2, 3]M.E Computer science and engineering, Prof. Ram Meghe Institute of technology & Research Badnera

Abstract: Today, Email Spam has become a major problem, with Rapid increament of internet users, Email spams is also increasing. People are using email spam for illegal and unethical conducts, phishing and fraud. Sending malicious link through spam emails which can damage the system and can also seek in into your system. Spammer creates a fake profile and email account which is easier for them. These spammers target those peoples who are not aware about frauds. So there is a need to identify the fraud in terms of spam emails. In this paper we will identify the spam by using machine learning algorithms.
Keywords: Spam; Spam filtering; Techniques; False positive;

I. INTRODUCTION

Email or electronic mail spam refers to the "using of email to send unsolicited emails or advertising emails to a group of recipients. Unsolicited emails mean the recipient has not granted permission for receiving those emails. "The popularity of using spam emails is increasing since last decade. Spam has become a big misfortune on the internet. Spam could be a waste of storage, time and message speed. Automatic email filtering could also be the foremost effective methodology of police investigation spam however today spammers will simply bypass of these spam filtering applications simply. Several years ago, most of the spam can be blocked manually coming from certain email addresses. Machine learning approach will be used for spam detection. Major approaches adopted closer to junk mail filtering encompass "text analysis, white and blacklists of domain names, and community-primarily based techniques". Text assessment of contents of mails is an extensively used method to the spams. Many answers deployable on server and purchaser aspects are available. Naive Bayes is one amongst the utmost well-known algorithms applied in these procedures. However, rejecting sends basically enthusiastic about content examination will be a troublesome issue within the event of counterfeit positives. frequently purchasers and organizations wouldn't would like any legitimate messages to be lost. The boycott approach has been probably the soonest technique pursued for the separating of spams. The technique is to acknowledge all the sends other than those from the area/electronic mail ids. Expressly boycotted. With more up to date areas coming into the classification of spamming space names this technique keeps an eye on no longer work so well. The white list approach is the approach of accepting the mails from the domain names/addresses openly white listed and place others in a much less importance queue, that is delivered most effectively after the sender responds to an affirmation request sent through the "junk mail filtering system".

II. MACHINE LEARNING IN EMAIL CLASSIFICATION[7]

Machine learning field may be a subfield from the broad field of AI, this aims to make machines able to learn like human. Learning here means understood, observe and represent information about some statistical phenomenon. In unsupervised learning one tries to uncover hidden regularities (clusters) or to observe anomalies within the information like spam messages or network intrusion. In e-mail filtering task some options may be the bag of words or the topic line analysis. Thus, the input to e-mail classification task is viewed.

A. Naive Bayes Classifier Model

Naïve Bayes[3] classifier was used in 1998 for spam recognition. The Naïve Bayes classifier algorithm is an algorithm which is used for supervised learning. The Bayesian classifier works on the dependent events and works on the probability of the event which is going to occur in the future that can be detected from the same event which occurred previously. Naïve Bayes was made on the Bayes theorem which assumes that features are autonomous of each other. Naïve Bayes classifier technique can be used for classifying spam emails as word probability plays main role here. If there is any word which occurs often in spam but not in ham, then that email is spam. Naive Bayes classifier algorithm has become a best technique for email filtering. For this the model is trained using the Naïve Bayes filter very well to work effectively. The Naive Bayes always calculates the probability of each class and the class having the maximum probability is then chosen as an output. Naïve Bayes always provide an accurate result. It is used in many fields like spam filtering.

©IJRASET: All Rights are Reserved

B. K-nearest Neighbour Classifier Model

K-Nearest Neighbors (KNN) is one of the simplest classification method used in Machine Learning for regression and classification problem. KNN algorithms use data and categorise new data points based on similarity measures (e.g. distance function).Classification is done by a majority vote to its neighbours. The data is assigned to the class which has the nearest neighbors. As you increase the number of nearest neighbours, the value of k, accuracy might increase.

C. Artificial Neural Network Classifier Model

An artificial neural network (ANN), is a computational model based on biological neural networks. It consists of an interconnected collection of artificial neurons. An artificial neural network is an adaptive system that changes its structure based on data that flows through the artificial network during a learning phase. There are, however, two classical kind of neural networks, perception and multilayer perception.

D. Support Vector Machine Classifier Model

SVM or Support Vector Machine[6] is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyper-plane which separates the data into classes.

III.IMPLEMENTATION AND RESULT ANALYSIS

A. Implementation

In this module, a dataset from "Kaggle" website is used as a training dataset. The inserted dataset is first checked for duplicates and null values for better performance of the machine. Then, the dataset is split into two sub-datasets; say "train dataset" and "test dataset" within the proportion of 70:30. Then for text-processing the "train" and "test" dataset is then passed as parameters. In text-processing, punctuation symbols and words that are within the stop words list are removed and came as clean words. For "Feature Transform" These clean words are then passed. In feature remodel the clean words that are came from the text-processing are then used for 'fit' and 'transform' to make a vocabulary for the machine. The dataset is additionally passed for "hyper parameter tuning" to seek out optimum values for the classifier to use in step with the dataset those values with a random state. The state of the trained model and features are saved for future use for testing unseen data.

B. Result Analysis

1) *Spam Email Detection By Existing:* We pass certain email content to both existing and proposed in which all mails are pass which are non-spam so below system show the detections of mails which are shown below TABLE I

TABLE I
SPAM EMAIL DETECTION BY EXISTING

Sr. No	Inputted Mail Contain	Is Spam detected?
1	Hi welcome to india we are very happy to see you here	No
2	Jubilant Foodworks just put out their quarterly results for March FY21 and they were pretty impressive, considering the pandemic. So we thought we'd take a look at the company and map their journey over the past few years.	No
3	Connect with a Reply These top videos on Twitter are the most played for a reason. Take a look, and then join the conversation about them by tapping Reply.	Yes

©IJRASET: All Rights are Reserved

4	Writing to tell you LeadSquared turns 10 this June and truly what a decade it has been! We were born with the single objective of enabling sales teams to drive high velocity sales by enabling higher productivity levels and ultimately more revenue. Over the decade we've become strategic sales advisors to over 2000+ clients and 390000 sales people. Leveraging automation, pre-emptive nudges, faster TATs and simple workflows - all in a DIY platform . Since going live, these are some of the metrics we are most proud of : 1. Achieve a DAU rate of approx 99%. 2. 2.5X increment in meetings completed per agent. 3. Drive 30% increase in revenues , earned within 2 quarters. 4. Increase in lead engagement by 2X. 5. Intelligent personal sales assistant and automation to help with a 30% improvement in productivity levels. 6. Real time analytics across hierarchies to help proactive jump ins for deals at risk. All of this in one single, cloud solution! Zero dependencies and delayed handoffs guaranteed. Given that this formulae is highly replicable, I am checking in to see if we can explore any way we could work together. I think this would be of tremendous value to your teams. You can book your schedule here with our inhouse Industry experts. Do let me know if sometime this week or next works for you? Excited to talk!	No
5	Get more of the good stuff To get the Tweets you love, just follow the accounts who bring great content. For example, check out the latest from @TwitterVideo, where the top videos of the week are served up fresh. If you like what you see, click the follow button to get the latest delivered right to your timeline. See more on Twitter	No
6	This email and any files transmitted with it are confidential and intended solely for the use of the individual or entity to whom they are addressed. If you have received this email in error please notify the system manager. This message contains confidential information and is intended only for the individual named. If you are not the named addressee you should not disseminate, distribute or copy this e-mail. Please notify the sender immediately by e-mail if you have received this e-mail by mistake and delete this e-mail from your system. If you are not the intended recipient you are notified that disclosing, copying, distributing or taking any action in reliance on the contents of this information is strictly prohibited . The information contained in this mail is propriety and strictly confidential.	No

©IJRASET: All Rights are Reserved

	The policy of your Maruti Wagon R expires tomorrow and we have just the right deal for you! Avoid paying hefty penalties for not having valid insurance. Renew now For any other assistance, please visit the support section on our website: Help & Support. Happy driving! Team Acko	
7	Discord makes communication more engaging and fun! It is known for building communities and letting them communicate using voice calls, text messages, and more. Though it was built for the gaming community, businesses are quickly realizing the platform's potential for marketing and more. The true magic of Discord, however, begins when you magnify its scope and potential by integrating other applications with it. How, you ask? Appy Pie Connect - the unique DIY workflow automation platform offers numerous meaningful integrations between Discord and hundreds of other valuable applications. Here are some of the most popular Discord integrations on Appy Pie Connect. Popular Discord integrations By connecting Discord and Gmail, you can automate multiple actions and save time and money. For example, you can send a Discord channel message automatically every time you receive a new Gmail email.	No
8	Greetings from Nippon India Mutual Fund! In compliance with the Go Green Initiative by SEBI, kindly click here to access the Portfolio of the schemes of Nippon India Mutual Fund for the month of May, 2021. This initiative is taken in pursuance to SEBI Circular No.SEBI/HO/IMD/DF2/CIR/P/2018/92 dated June 05, 2018, on disclosing and providing information to unit holders.	Yes

Total Numbers of Email Received =8
Total SPAM Email Detected = 2
Recall Calculated =6/8=0.75
Precision : 2/8=0.25

©IJRASET: All Rights are Reserved

2) Spam Email Detection By proposed

TABLE II
SPAM EMAIL DETECTION BY PROPOSED

Sr. No	Inputted Mail Contain	Is Spam detected?
1	Hi welcome to india we are very happy to see you here	No
2	*Jubilant Foodworks just put out their quarterly results for March FY21 and they were pretty impressive, considering the pandemic. So we thought we'd take a look at the company and map their journey over the past few years.*	No
3	**Connect with a Reply** These top videos on Twitter are the most played for a reason. Take a look, and then join the conversation about them by tapping Reply.	No
4	Writing to tell you LeadSquared turns 10 this June and truly what a decade it has been! We were born with the single objective of enabling sales teams to drive high velocity sales by enabling higher productivity levels and ultimately more revenue. Over the decade we've become strategic sales advisors to over 2000+ clients and 390000 sales people. Leveraging automation, pre-emptive nudges, faster TATs and simple workflows - all in a DIY platform . Since going live, these are some of the metrics we are most proud of : 1. Achieve a DAU rate of approx 99%. 2. 2.5X increment in meetings completed per agent. 3. Drive 30% increase in revenues , earned within 2 quarters. 4. Increase in lead engagement by 2X. 5. Intelligent personal sales assistant and automation to help with a 30% improvement in productivity levels. 6. Real time analytics across hierarchies to help proactive jump ins for deals at risk. All of this in one single, cloud solution! Zero dependencies and delayed handoffs guaranteed. Given that this formulae is highly replicable, I am checking in to see if we can explore any way we could work together. I think this would be of tremendous value to your teams. You can book your schedule here with our inhouse Industry experts. Do let me know if sometime this week or next works for you? Excited to talk!	No

©IJRASET: All Rights are Reserved

5	**Get more of the good stuff** To get the Tweets you love, just follow the accounts who bring great content. For example, check out the latest from @TwitterVideo, where the top videos of the week are served up fresh. If you like what you see, click the follow button to get the latest delivered right to your timeline. **See more on Twitter**	No
6	This email and any files transmitted with it are confidential and intended solely for the use of the individual or entity to whom they are addressed. If you have received this email in error please notify the system manager. This message contains confidential information and is intended only for the individual named. If you are not the named addressee you should not disseminate, distribute or copy this e-mail. Please notify the sender immediately by e-mail if you have received this e-mail by mistake and delete this e-mail from your system. If you are not the intended recipient you are notified that disclosing, copying, distributing or taking any action in reliance on the contents of this information is strictly prohibited . The information contained in this mail is propriety and strictly confidential. The policy of your Maruti Wagon R expires tomorrow and we have just the right deal for you! Avoid paying hefty penalties for not having valid insurance. Renew now For any other assistance, please visit the support section on our website: Help & Support. **Happy driving!** Team Acko	No
7	Discord makes communication more engaging and fun! It is known for building communities and letting them communicate using voice calls, text messages, and more. Though it was built for the gaming community, businesses are quickly realizing the platform's potential for marketing and more. The true magic of Discord, however, begins when you magnify its scope and potential by integrating other applications with it. How, you ask? Appy Pie Connect - the unique DIY workflow automation platform offers numerous meaningful integrations between Discord and hundreds of other valuable applications. Here are some of the most popular Discord integrations on Appy Pie Connect. **Popular Discord integrations**	No

©IJRASET: All Rights are Reserved

	By connecting Discord and Gmail, you can automate multiple actions and save time and money. For example, you can send a Discord channel message automatically every time you receive a new Gmail email.	
8	Greetings from Nippon India Mutual Fund! In compliance with the Go Green Initiative by SEBI, kindly click here to access the Portfolio of the schemes of Nippon India Mutual Fund for the month of May, 2021. This initiative is taken in pursuance to SEBI Circular No.SEBI/HO/IMD/DF2/CIR/P/2018/92 dated June 05, 2018, on disclosing and providing information to unitholders.	Yes

Total Numbers of Email Received =8
Total SPAM Email Detected = 1
Recall Calculated =7/8=0.87
Precision:1/8=0.12

IV. CONCLUSIONS

With this result, Naïve Bayes gives the best outcome but has limitation due to class-conditional independence which makes the machine to misclassify some tuples. Nowadays, lots of emails are sent and received and it is difficult as our project is only able to test emails using a limited amount of corpus. Thus spam detection is proficient of filtering mails giving to the content of the email and not according to the domain names or any other criteria. Therefore, at this it is an only limited body of the email. There is a wide possibility of improvement in our project. The subsequent improvements can be done: “Filtering of spams can be done on the basis of the trusted and verified domain names.” “The spam email classification is very significant in categorizing e-mails and to distinct e-mails that are spam or non-spam.” “This method can be used by the big body to differentiate decent mails that are only the emails they wish to obtain.”

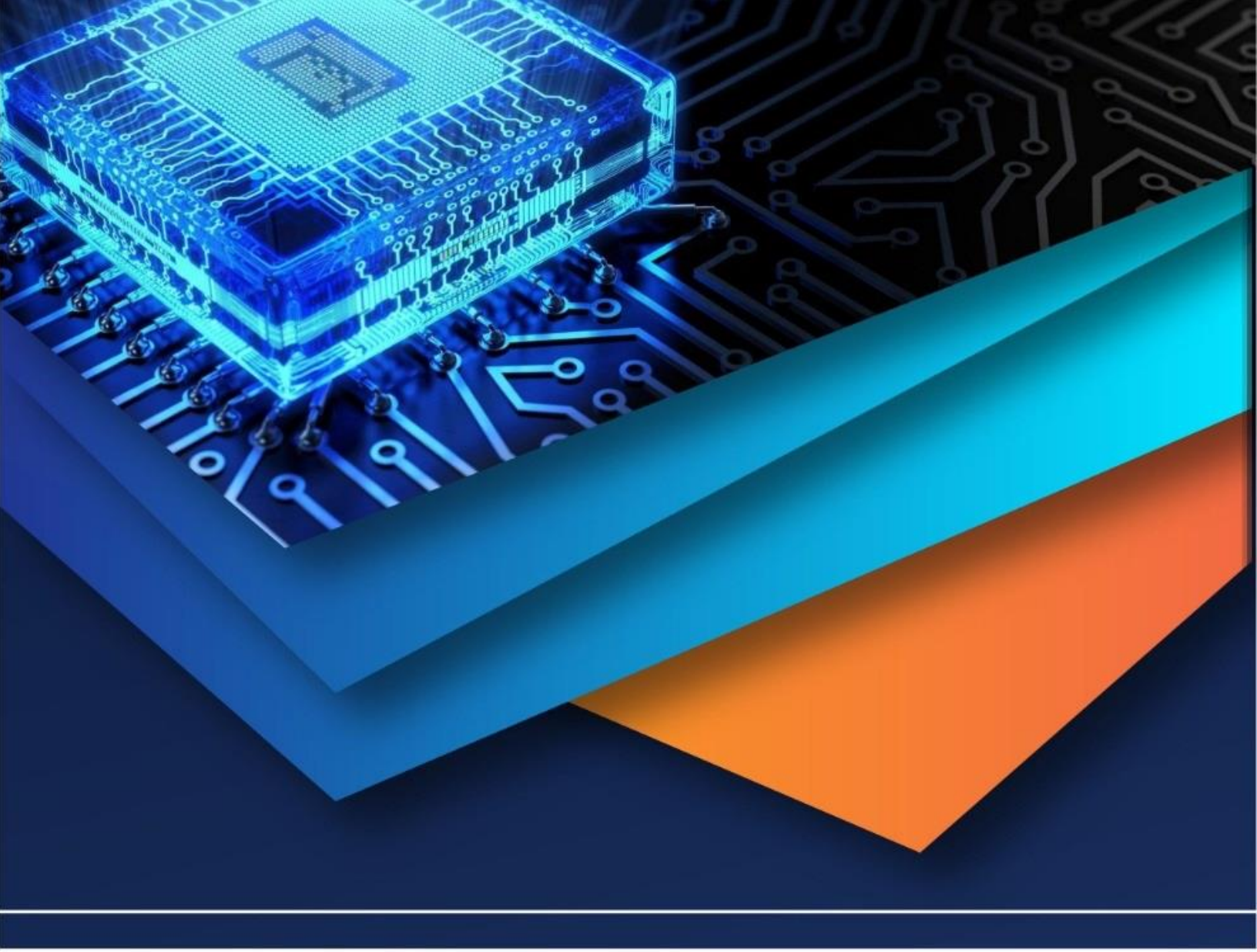
V. ACKNOWLEDGMENT

It is our proud privilege to release the feelings of our gratitude to every person who helped us directly or indirectly to conduct this research work. we express our heart full indebtness and owe a deep sense of gratitude to Prof. Sangram Dande for their sincere guidance and inspiration for completing this paper

REFERENCES

[1] Suryawanshi, Shubhangi & Goswami, Anurag & Patil, Pramod.(2019). Email Spam Detection: An Empirical Comparative Study of Different ML and Ensemble Classifiers. 69-74. 10.1109/IACC48062.2019.8971582.
[2] Karim, A., Azam, S., Shanmugam, B., Krishnan, K., & Alazab, M. (2019). A Comprehensive Survey for Intelligent Spam Email Detection. IEEE Access, 7, 168261-168295. [08907831]. https://doi.org/10.1109/ACCESS.2019.2954791
[3] K. Agarwal and T. Kumar, "Email Spam Detection Using Integrated Approach of Naïve Bayes and Particle Swarm Optimization," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 685-690.
[4] Harisinghaney, Anirudh, Aman Dixit, Saurabh Gupta, and Anuja Arora. "Text and image-based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm." In Optimization, Reliabilty, and Information Technology (ICROIT), 2014 International Conference on, pp.153-155. IEEE, 2014
[5] Mohamad, Masurah, and Ali Selamat. "An evaluation on the efficiency of hybrid feature selection in spam email classification." In Computer, Communications, and Control Technology (I4CT), 2015 International Conference on, pp. 227-231. IEEE, 2015
[6] Shradhanjali, Prof. Toran Verma “E-Mail Spam Detection and Classification Using SVM and Feature Extraction”in International Jouranl Of Advance Reasearch, Ideas and Innovation In Technology,2017 ISSN: 2454-132X Impact factor: 4.295
[7] W.A, Awad & S.M, ELseuofi. (2011). Machine Learning Methods for Spam E-Mail Classification. International Journal of Computer Science & Information Technology. 3. 10.5121/ijcsit.2011.3112.
[8] A. K. Ameen and B. Kaya, "Spam detection in online social networks by deep learning," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1-4.
[9] Diren, D.D., Boran, S., Selvi, I.H., & Hatipoglu, T. (2019). Root Cause Detection with an Ensemble Machine Learning Approach in the Multivariate Manufacturing Process.
[10] Tasnim Kabir, Abida Sanjana Shemonti, Atif Hasan Rahman. "Notice of Violation of IEEE Publication Principles: Species Identification Using Partial DNA Sequence: A Machine Learning Approach”, 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE), 2018.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)