iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** V **Month of publication:** May 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.53235

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 11 Issue V May 2023- Available at www.ijraset.com*

# Spam SMS Prediction Using Machine Learning

Esha Choudhary[1], Bhanu Verma[2], Aakhya Chaudhary[3], Anushka Sengar[4], Ayushi Agarwal[5]
*Department of Computer Science, IMS Engineering College,Ghaziabad, Uttar Pradesh, India*

***Abstract:*** ***As the popularity of mobile phone devices has increased, Short Message Service (SMS) has grown into a multi-billion dollars industry. At the same time, reduction in the cost of messaging services has resulted in growth in unsolicited commercial advertisements (spams) being sent to mobile phones. In parts of Asia, up to 30% of text messages were spam in 2012. Lack of real databases for SMS spams, short length of messages and limited features, and their informal language are the factors that may cause the established email filtering algorithms to underperform in their classification. In this project, a dataset of real SMS Spams from UCI Machine Learning repository is used, and after pre-processing and vectorization, different machine learning algorithms are applied to the dataset. Finally, the results are compared and the best algorithm for spam filtering for text messaging is introduced and converted into an open-source website. The SMS spam collection set is used for testing the method. After collecting the various supervised learning algorithms, we find that the Multinomial Naïve Bayes algorithm gives us 97.1% Accuracy and 100% Precision.***

***Keywords: SMS Spam Classification, Machine Learning, Multinomial Naïve Bayes, Supervised Learning***

## I. INTRODUCTION

### *A. What is Spam and why should it be Prevented?*

Spam is unsolicited and unwanted messages sent electronically and whose content may be malicious. Email spam issent/received over the Internet while SMS spam is typically transmitted over a mobile network. We'll refer to user that sent spam as 'spammers'. SMS messages are usually very cheap (if not free) for the user to send, making it appealing for unrightful exploitation. This is further aggravated by the fact that SMS is usually regarded by the user as a safer, more trustworthy form of communication than other sources, e. g., emails. .The origin of the term "spam" for invasive bulk messaging refers to a Monty Python Skit .

The dangers of spam messages for the users are many: undesired advertisement, exposure of private information, becoming a victim of a fraud or financial scheme, being lured into malware and phishing websites, involuntary exposition to inappropriate content, etc. For the network operator, spam messages result in an increased cost in operations. Spam SMS are unwanted messages sent by an anonymous number which may have malicious content. The users who send these kind of SMS's are referred to as spammers. Nowadays people are using mobile phones so much that the spread of Spam SMS is also increasing. There are so many dangers in spam SMS like: undesired advertisement, exposure of private information, becoming a victim of a fraud or financial scheme, being lured into malware and phishing websites, involuntary exposition to inappropriate content, etc. Spam is common on social media sites like YouTube, and it mainly consists of comments and links to pornographic websites, as well as irrelevant videos. These comments are sometimes created automatically by bots. Although the definition of spam on online video game sharing services is debatable, instances of message flooding, requests to join a specific group, violations of copyrights, and so on are occasionally referred to as spam. Spam in blogs, often known as splog, refers to comments that have nothing to do with the topic of discussion. Frequently, these comments are accompanied by links to commercial websites. Some splogs are devoid of unique content and contain stuff plagiarized from other websites. Spammers use multiple numbers in order to send these kind of harmful messages so number blocking is not enough to stop spammers. That's why spam filtering is required that relies not only on volume but also on the content of the SMS itself. Nowadays more than ever Spam SMS are flying around. It really doesn't matter what phone you own and where are you living spam messages will come in your phone and try to con you out of money. They can be convincing or not-at-all convincing but they are creating interruption in people's lives. A company named Robo Killer whose researches shows the SMS Spam in U.S is increased by 28% between February and March alone this year. The company haven't seen this much increase in SMS Spam since 2017.the purpose of SMS Spam can be marketing and announcement of a variety of products ,money theft, political issues, identity theft. SMS Spam Classification is an significant task in which SMS are classified into Spam and Ham. SMS Flooding is a very serious problem.

In the case under study, spam is an annoyance to the user and thus detrimental to the quality of the service that hurts the brand in the process. This can lead to complaints, low ratings and even loss of users, not to mention users getting scammed.

*B. Differences with Email Spam*

Table 1: The main differences between Spam in Email and SMS

| SMS | EMAIL |
|---|---|
| Short messages | Can have any length |
| Sent (mostly) through mobile connections | Transmitted through any internet connection |
| Ambiguous intention | Greater length makes the intention clearer |
| Content can be plain text, string characters and possibly emojis | Has a subject, formatted text, multimedia content and attachments |
| Usually regarded as trustworthy | There is widespread awareness about spam emails |

Spam emails are also known as junk emails; these emails are unsolicited messages which are bulk and sent with an expectation of getting a small amount of interaction.
With the increase in technology , Spam has also increased tremendously. No matter how much new channels are getting discovered for marketing and communication, email always holds the first spot.
Short message service(SMS) enables a mobile device to send, receive and display messages of up to 160 characters in Roman text and variations for non-Roman character sets. Messages received are stored in the network if the subscriber device is inactive and are relayed when it becomes active. SMS has become available increasingly in CDMA networks and in some fixed network.

*C. Spammers' Behavior*

Spammers use many forms of communication to bulk-send their unwanted messages. Some of these are marketing messages peddling unsolicited goods. Other types of spam messages can spread malware, trick you into divulging personal information, or scare you into thinking you need to pay to get out of trouble. Spammers attempt to test the operator's anti-spam infrastructure by to send messages, which rules out number blocking as an strategy to prevent spam. This situation requires some kind of content-based filtering that relies not only on volume but also on the content of the SMS itself.

*D. Difficulties*

*1)* There are no publicly available large datasets of spam SMS. Even if there were, it is not at all expected that training on those datasets would translate into a good performance within our context. Therefore, there is no choice but to build a custom dataset from real data that streams into the system.
*2)* Absence of a pipeline to transform SMS logs into a structured and clean dataset.
*3)* The app is available in many countries and languages, adding another layer of complexity.
*4)* The ultimate model has to be deployed and integrated within the current infrastructure of the app taking the necessary precautions to avoid incurring high costs and delays in message delivery.
*5)* Subjective criteria for labeling: is it okay to block religious propaganda, even if it is not intended as a fraud or scam? And what if the message is broadcast to thousands of users?
*6)* Message ambiguity: it can even be hard for a human to distinguish between real messages and spam.

Examples
These are examples based on real messages (not all were spam):
*a)* Your package is awaiting delivery bit.ly/xxxxxxx
*b)* How fast are your fingers? Test Now! -> https://play.google.com/store/apps/details?idxxxxxxx)
*c)* Good evening sir, happy resumption. I had called your number today but you did not pick it. Please forward watchword money to my account.

- I use PayTree sent you up to € 25 Sign up with my link to claim, then get €300, 000 give away fund: https://palmpay.site/QXAfLgKsPhKA
- You have received USD150 from John Carpenter (+1)11111111 in your Mobile Money account on 2021–04–22 Message from sender: Transaction ID: 2901380912. For a successful cash out Ecobank will contact you for more inquiries.

- Follow this link to join my WhatsApp group: http://unnoficcial_whatsapp.com/download
- Please check and send to me, my result 100 level third year . Name: John-Paul Gillian Jambaya Email: johnpaulGillianJ@gmail.com Jamb reg no. 49851407SA Metric Number. YU/14/6337 Password: 2352246811678. I will pay you.

## II. LITERATURE REVIEW

### A. *Related Work*

After the study of SMS Spam detection the researches include various strategies, such as Bayesian based classifiers,logistic regression, support vector machine, decision tree strategy.E-mail classification has been an active area of research. Cohen (1996) developed a propositional learning algorithm RIPPER to induce ‘‘keyword-spotting rules’’ for filing e-mails into folders. The multi-class problem was transformed into several binary problems by considering one folder versus all the others.Cohen argued that keyword spotting rules are more useful as they are easier to understand, modify and can be used in conjunction with user-constructed rules. Sahami (1998) applied NB for spam e-mail filtering using bag of words to represent e-mail corpora and binary encoding. The performance improved by the incorporation of hand-crafted phrases and domain-specific features such as the domain type of the sender and the percentage of non-alphabetical characters in the subject. Rennie (2000) used Naïve-Bayes to filee-mails into folders and suggested the three most suitable folders for each message. The system applies stemming, removes stop words and uses document frequency threshold as feature selector.Pantel et al. (1998) developed SpamCop: a spam classification and organization program. SpamCop is a system for spam e-mail filtering also based on Naïve-Bayes. Both stemming and a dynamically created stopword list are used. The authors investigated the implications of the training data size, different rations of spam and non-spam e-mails, use of trigrams instead of words and also showed that Spam Cop out performed Ripper. MailCat et al. (1999) uses a nearest-neighbor (k-NN) technique and tf-idf representation to filee-mails into folders. KNN supports incremental learning but requires significant time for classification of new e-mails. Androutsopoulos et al. (2000) found that Naïve-Bayes and a k-NN technique called TiMBLclearly outperform the keyword-based spam filter of Outlook 2000 on the Ling Spam corpora. Ensembles of classifiers were also used for spam filtering. Sakkis et al. (2001) combined a NB and k-NN by stacking and found that the ensemble achieved better performance. Carrera & Marquez (2001) showed that boosted trees outperformed decision trees, NB and k-NN. Rios &Zha (2004) applied RF for spam detection on time indexed data using a combination of text and metadata features. For low false positive spam rates, RF was shown to be overall comparable with SVM in classification accuracy.

### B. *Different Researcher’s Contribution*

Table 2: It discusses the different contributions in Spam Filtering and the techniques

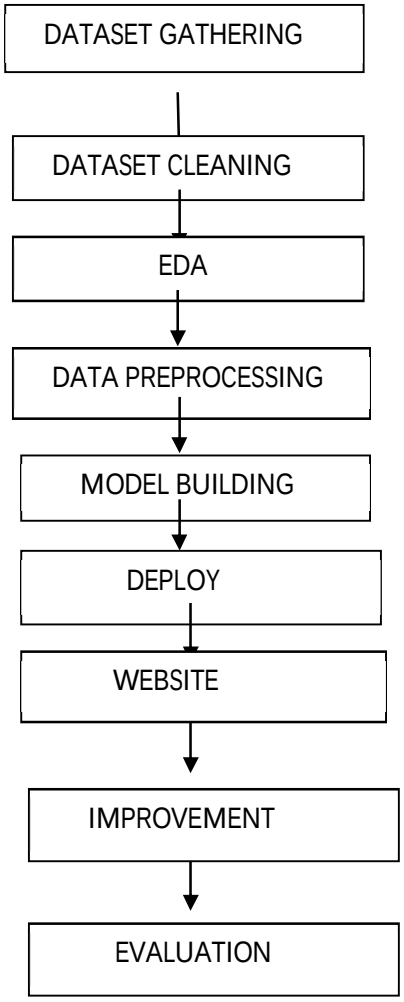
| Authors | Year | Technique |
|---|---|---|
| Gomatham Sai Sravya, G Pradeepini, Vaddeswaram, Guntur[1] | 2020 | MobileSms Spam Filter Techniques using Machine Learning Techniques |
| PavasNavaney ,GauravDubey,Ajay Rana[2] | 2018 | SMS Spam Filtering Using Supervised Machine Learning Algorithms |
| Suparna Das Gupta ,SoumyabrataSaha, Suman Kumar Das[3] | 2020 | SMS Spam Detection using Machine Learning |
| Nilam Nur Amir Sjarif, Nurulhuda Firdaus Mohd Azmi, SuriayatiChuprat, Haslina Md Sarkan, Yazriwati Yahya, SurianiMohdSam[4] | 2019 | SMS Spam Message Detection using Term Frequency –Inverse Document Frequency and Randoom Forest Algorithm |
| Nikunj Chaudhari, Prof. Jayvala, Prof. Vinitashah[5] | 2016 | Survey on SMS Filtering using Data Mining Techniques |
| SamadhanNagre[6] | 2018 | Mobile SMS Spam Detection Using Machine Learning Techniques |
| Abhishek Patel, Priya Jhariya, SudalaguntaBharath, Ankita wadhawan[7] | 2021 | SMS Spam Detection using Machine Learning Approach . |

| | | |
|---|---|---|
| Kavya P, Dr. A. Rengarajan[8] | 2020 | A Comparative Study For SMS Spam Detection |
| Olusola Abayomi-Alli, Sanjay Misra,Adebayo Abayomi-Alli[19] | 2022 | A Deep Learning Method For automatic SMS Spam Classification :Performance of learning Algorithms on Indigenous dataset |
| Muhammad Adeel Abid, Saleem Ullah, Muhammad Abubakar Siddique, [20] | 2022 | Spam SMS Filtering based on text features and supervised machine learning techniques |
| Luo GuangJun,[1]Shah Nazir,[2]Habib Ullah Khan,[3]and Amin Ul Haq[4] [21] | 2020 | Spam Detection approach for secure mobile message communication using machine learning algorithm |

## III. DESCRIPTION OF WORK

The dataset used is gathered by UCI Machine Learning Repository it has total 5572 messages from which 4825 are Ham and 747 are Spam. The cleaning of data and preprocessing is performed so that the dataset will be perfect for model building. Exploratory data analysis is also performed to analyse the trends and patterns in our dataset. After study of Research papers mentioned in Table 1 different machine learning algorithms are used and then compared to find out which technique gives most accurate and précised results. The different techniques used are Naïve Bayes Classifier, Logistic Regression, Support Vector Classifier, Decision Tree, K-Nearest Neighbour , Random Forest Classifier, AdaBoost Classifier, Bagging Classifier, Gradient Boosting Algorithm and XGBoost Algorithm, The Voting classifier is also used by combining various best performing algorithms. For vectorization of dataset two techniques are used Bag of Words and TF-IDF. After building the model a pipeline will be created for making a website by using streamlit framework.

### *A. Flowchart*



*B. PSEUDO Code Of Proposed System*

Table 3: The steps performed in this work

| Steps | Overview |
|---|---|
| Step 1 | Import dataset and then perform data cleaning steps. |
| Step 2 | Perform EDA for understanding the dataset |
| Step 3 | Perform data pre-processing steps |
| Step 4 | Train the model by using naïve bayes classifier |
| Step 5 | Also compare accuracy and precision of different machine learning algorithms. |
| Step 6 | After training model, persist model by saving .pkl file for using model in future without refrain |
| Step 7 | Make website by using Streamlit |
| Step 8 | Delpoy it on Heroku platform |

*C. Methodology used*

*1) Spam Filtering Process*

SMS are classified into Spam and not spam by giving them as input to various algorithms by using steps:

*a)* *Data Cleaning:* Data cleaning is performed to check if there are any unnamed rows in our data, if data contains any null values and duplicate values.

*b)* *Exploratory Data Analysis:* EDA is performed to understand the data as if our dataset is balanced or not,to check how many spam and ham messages are present in our data and to find the ratio of ham and spam SMS.

*c)* *Data pre-processsing:* Data pre-processing is used to convert data into predictable format. In this stage stop words are eliminated, stemming and tokenization is performed.

*d)* *Vectorization:* Before model building data is vectorized.After comparing the results of Bag of words and TF-IDF vectorization, TF-IDF is used.

*e)* *Model Building:* After comparing different machine learning techniques Multinomial Naïve Bayes Algorithm is used for building the model.

*f)* *Improving Model:* Tried to improve model by changing max feature of TF-IDF ,usingVoting classifier and stacking and scaling. Only by changing the max feature of TF-IDF the accuracy of Naïve bayes is increased so it was used in the model.

*2) Algorithms*

*a)* *Naïve Bayes Classifier:* Gaussian Naïve Bayes, Multinomial Naïve Bayes and Bernoulli naïve algorithms are used for SMS Spam Detection. Multinomial Naïve Bayes: It assigns documents to classes on the basis of statistical analysis of their contents. It is very easy to use on continuous and discrete data. It has ability to handle large datasets. For training Natural Language processing models ,it is best.

```
mb.fit(x_train,y_train)
y_pred2 = mnb.predict(x_test)
print(accuracy_score(y_test,y_pred2))
print(confusion_matrix(y_test,y_pred2))
print(precision_score(y_test,y_pred2))
//we want more and more precision score
0.9709864603481625
[ [896 0]
[30 108] ]
1.0
```

Fig 1. Output of Multinomial Naïve Bayes algorithm

Multinomial Naïve bayes algorithm provided 97% accuracy score and 100% precision score so it is used with TF-IDF vectorizer. Different algorithms are also used to check if they give more accuracy then Multinomial naïve bayes.

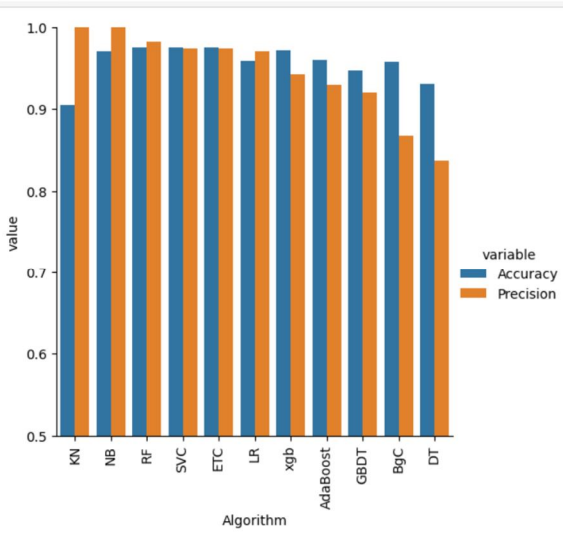



Fig 2. Comparision between different algorithms used

As shown in figure 3 K-Nearest Neighbour algorithm and Naïve Bayes is performimg best. Precision score of K-Nearest Neighbour is less than Naïve Bayes so it will not be used.
For improving the model different types of techniques are used like : 1) Changing the max feature of TF-IDF

| Algorithm | Accuracy | Precision | Accuracy_max_ft_3000 | Precision_max_ft_3000 |
|---|---|---|---|---|
| KN | 0.900387 | 1.000000 | 0.905222 | 1.000000 |
| NB | 0.959381 | 1.000000 | 0.971954 | 1.000000 |
| ETC | 0.977756 | 0.991453 | 0.979691 | 0.975610 |
| RF | 0.970019 | 0.990826 | 0.975822 | 0.982906 |
| SVC | 0.972921 | 0.974138 | 0.974855 | 0.974576 |
| AdaBoost | 0.962282 | 0.954128 | 0.961315 | 0.945455 |
| xgb | 0.971954 | 0.950413. | 0.968085 | 0.9333884 |
| LR | 0.951644 | 0.940000 | 0.956480 | 0.969697 |
| GBDT | 0.951644 | 0.931373 | 0.946809 | 0.927835 |
| BgC | 0.957447 | 0.861538 | 0.959381 | 0.869231 |
| DT | 0.935203 | 0.838095 | 0.931335 | 0.831683 |

Fig. 3 After changing Max Feature value the accuracy and precision score

By changing the max feature to 3000 accuracy of Naïve bayes is increased to 97.1% so it is used in the model.
Other Techniques are also used to check if it will improve the accuracy of Naïve Bayes like Scaling ,stacking and combining the algorithms but it does not improve the accuracy of Naïve Bayes . Hence, it can be concluded that multinomial naïve bayes can be used to classify spam and ham SMS with 97% Accuracy and 100% precision.

## IV. RESULTS AND DISCUSSIONS

The goal of this project is to classify the Spam and not Spam SMS and then converting it into a website using streamlit framework. SMS SPAM is any junk message which is sent to a person containing malicious content which can lead to identity theft, money theft and cyber attacks. The dataset for this project originates from the UCI Machine Learning Repository. The different techniques are used to differentiate between spam and ham SMS like primarily based on the length of message, unique keywords etc. Accuracy and precision of various machine learning algorithms are compared.

Total Dataset: 5572 HAM SMS: 747 SPAM SMS: 747

Table 4: Accuracy and Precision Score of Algorithms used

| S.NO | Algorithm | Accuracy | Precision |
|---|---|---|---|
| 1. | Multinomial Naive Bayes | 97.1% | 100% |
| 2. | K-Nearest Neighbour | 90.5% | 100% |
| 3. | Extra Trees Classifier | 97.6% | 97.5% |
| 4. | Random Forest | 97.5% | 98.2% |
| 5. | Support Vector Classifier | 97.4% | 97.4% |
| 6. | AdaBoost Classifier | 96.1% | 94.5% |
| 7. | XGBoost | 95% | 96.8% |
| 8. | Gradient Boosting | 94.6% | 92.7% |
| 9. | Bagging Classifier | 95.9% | 86.9% |
| 10. | Decision Tree | 93.1% | 83% |

By comparing different algorithms , it shows that Multinomial Naïve Bayes using TF-IDF Vectorizer gives the better accuracy with 97.1% and 100% precision.

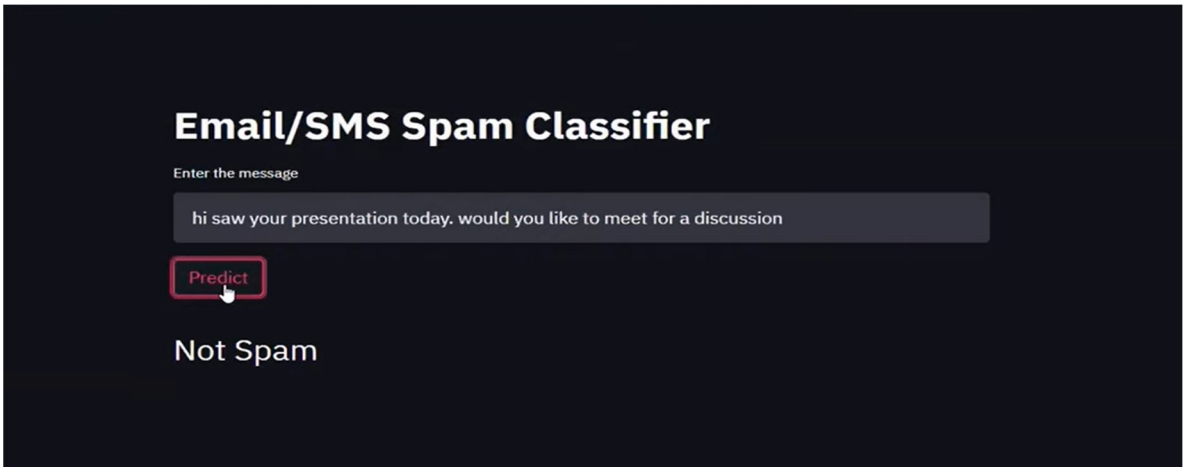



Fig. 4 Snapshot

This is the final output of project.

Overall it was possible to achieve a model able to identify spam reasonably well, without penalizing normal messages. The model seems to be generalizing to patterns not seen in the training set. Simple cases of spam in which only a number or a word changes with respect to a spam message already seen before are easily detected.

## V. CONCLUSION AND FUTURE SCOPE

From the above discussion and results it is concluded that Machine learning techniques play a vital role in classifying Spam and Non Spam SMS. The Accuracy obtained in this work is 97.1% by using Multinomial Naïve Bayes Algorithm. Detection of spam is important for securing message and e-mail communication. The accurate detection of spam is a big issue, and many detection methods have been proposed by various researchers. However, these methods have a lack of capability to detect the spam accurately and efficiently.

To solve this issue, we have proposed a method for spam detection using machine learning predictive models. The method is applied for the purpose of detection of spam. The experimental results obtained show that the proposed method has a high capability to detect spam. The proposed method achieved 99% accuracy which is high as compared with the other existing methods. Thus, the results suggest that the proposed method is more reliable for accurate and on-time detection of spam, and it will secure the communication systems of messages and e-mails. In this paper the existing spam detection techniques, their current applications and their limitations have been highlighted. It has been seen that even the existing techniques consist of certain loop holes and none of the methods is completely effective in itself. Reviewing the spam detection is indeed a hard task but it requires continuous research and development in this field.Any Machine learning project has room for improvement and this one is no different. These are some possible in order of predicted impact.

*A. Within the Current Framework*

*1)* Improve labels. This is by far the most important thing. The quality of the data is very highly correlated with thequality of the predictions. Some of the possibilities are:

*a)* Use message clustering after feature extraction to group similar messages together and manually label a representative of the group. The label is then shared to all data points in the group.

*b)* Come up with better labeling heuristics.

*c)* Leverage outside data such as sender, carrier, location, etc., to recognize spammers and label all their messages as spam.

*d)* Use spam data from other sources (with care) if it helps training.

*e)* More manual labeling.

*2)* Try different model architectures and tune hyper-parameters.

*3)* Model calibration so that spam probabilities actually represent a mathematically defined probability.

*B. Outside the Current Framework*

Including information besides the content of the SMS could provide supplementary variables to help decide which action to take on an SMS after its run through the model. User metadata such as tenure with the app, message frequency, count of previously blocked messages, etc… should be a great complement to spam probabilities computed solely from SMS content.

An integrated framework in which Spam probability is just one component will enhance the model's usefulness by reducing the number of false positives (from only blocking after a certain amount of warnings), not evaluating messages from known legitimate users and promptly block repeated spammers from using the app.

## VI. ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to Prof. (Dr.) Sonia Juneja, HoD in Department of Computer Science, IMSEC, Ghaziabad, for her stimulating guidance, continuous encouragement and supervision throughout the course of present work.

We are extremely thankful to Prof. (Dr.) Vikram Bali, Director, IMSEC, Ghaziabad, for providing us infrastructural facilities to work in, without which this work would not have been possible.

We would also like to express our genuine gratitude to Project Coordinator Mr. Awdhesh Kumar for his valuable suggestions and advices in carrying out this work.

We would like to thank our mentor Ms. Bhanu Verma, who have been our constant support in this project and helped us throughout the journey.

We would also like to thank the entire institute faculty who helped us directly or indirectly in completing our work.

## REFERENCES

[1] Gomatham Sai Sravya, G Pradeepini, Vaddeswaram, Guntur (2020), " MobileSms Spam Filter Techniques Using Machine Learning Techniques"

[2] PavasNavaney ,GauravDubey,Ajay Rana(2018) ,"SMS Spam Filtering Using Supervised Machine Learning Algorithms"

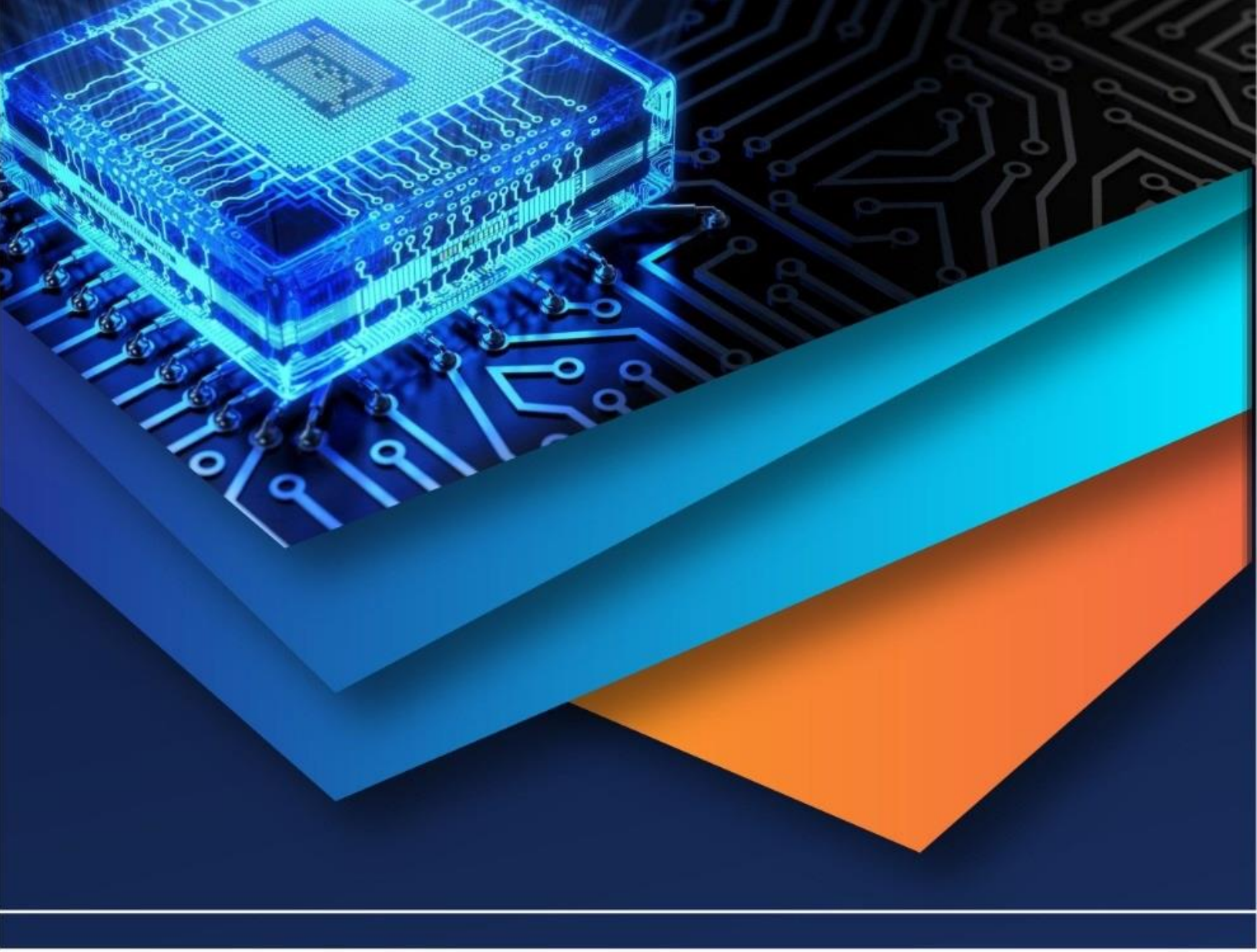
[3] Suparna Das Gupta ,SoumyabrataSaha, Suman Kumar Das (2020) ,"SMS Spam Detection using Machine Learning", Department of Information Technology, JIS College of Engineering, Zensar Technologies ,Pune,411014, Maharashtra, India

[4] Nilam Nur Amir Sjarif, Nurulhuda Firdaus Mohd Azmi, SuriayatiChuprat, Haslina Md Sarkan, Yazriwati Yahya, SurianiMohdSam(2019) ,"SMS Spam Message Detection using Term Frequency- Inverse Document Frequency and Random Forest Algorithm" , Razak Faculty of Technology and Informatics, UniversitiTeknologi Malaysia Kuala Lumpur, Level 5, Menara Razak, 54100 Jalan Sultan Yahya Petra, Kuala Lumpur, Malaysia

[5] Nikunj Chaudhari, Prof. Jayvala, Prof. Vinitashah, "Survey on SMS filtering using Data Mining Techniques", PG Scholar, Department of IT, G.H Patel College of Engineering and Technology, V.V Nagar, Anand,Assistant Professor, Department of IT, G.H Patel College of Engineering and Technology, V.V Nagar, Anand

[6] Samadhan Nagre (2018), "Mobile SMS Spam Detection Using Machine Learning Techniques", Dept of Computer Science and IT, Dr. B.A.M. University of Aurangabad
[7] Abhishek Patel, Priya Jhariya, SudalaguntaBharath, Ankita wadhawan, "SMS Spam Detection using Machine Learning Approach", Computer Science Engineering, Lovely Professional university, Phagwara, Punjab
[8] Kavya P, Dr. A. Rengarajan , "A Comparative Study for SMS Spam Detection", Master of Computer Application,Jain Deemed to be University, Bengaluru, Karnataka, India, School of CS and IT, Jain to be Deemed University, Bengaluru ,Karnataka ,India
[9] I. Alsmadi and I. Alhami, "Clustering and classification of email contents," Journal of King Saud University—Computer and Information Sciences
[10] B. Yu and Z.-B. Xu, "A comparative study for content-based dynamic spam classification using four machine learning algorithms," Knowledge-Based Systems
[11] A. Sharaff, "Comparative study of classification algorithms for spam email detection," in Emerging Research in Computing, Information, Communication and Applications
[12] S. Youn and D. McLeod, "A comparative study for email classification," in Advances and Innovations in Systems, Computing Sciences and Software Engineering
[13] R. K. Kumar, "Comparative study on email spam classifier using data mining techniques," in Proceedings of the International MultiConference of Engineers and Computer Scientists
[14] S. K. Trivedi and S. Dey, "Interplay between probabilistic classifiers and boosting algorithms for detecting complex unsolicited emails," Journal of Advances in Computer Networks
[15] P. K. Roy, J. P. Singh, and S. Banerjee, "Deep learning to filter SMS Spam," Future Generation Computer Systems
[16] R. Kaur, S. Singh, and H. Kumar, "Rise of spam and compromised accounts in online social networks: a state-of-the-art review of different combating approaches," Journal of Network and Computer Applications
[17] ShebutiRayana, Leman Akoglu Stony, Collective Opinion Spam Detection : Bridging Review Networks and the Metadata, KDD 2015
[18] Somayeh Shojaee, MasrahAzrifah Azmi Murad, Azreen Bin Azman, NurfadhlinaMohdSharef and SamanehNadali, Detecting Deceptive Reviews using Lexical and Syntactic Features, IEEE 2013
[19] Olusola Abayomi-Alli, Sanjay Misra,Adebayo Abayomi-Alli, "A Deep Learning method for automatic SMS Spam classification: Performance of learning algorithms on indigenous dataset", Department of software engineering. Kaunas, University of technology,Kaunas,Lithuania
[20] Muhammad Adeel Abid, Saleem Ullah, Muhammad Abubakar Siddique,Muhammad Faheem Mushtaq, Wajdi Aljedaani, Furqan Rustam, " Spam SMS filtering based on text features and supervised machine learning techniques"
[21] **Luo GuangJun**,[1]Shah Nazir,[2]Habib Ullah Khan,[3]and Amin Ul Haq[4] , "Spam Detection approach for secure mobile message communication using Machine Learning Algorithms"
[22] Olubodunde Agboola , " Spam Detection Using Machine Learning and Deep Learning", Louisiana State University and Agricultural and Mechanical College
[23] C. Oswald, Sona Elza Simon ,Arnab Bhattacharya, "SpotSpam: Intention Analysis-driven SMS Spam Detection using Bert Embeddings
[24] Pradeep Kumar Roy, Jyoti Prakash Singh, Snehasish Banerjee, "Deep learning to filter SMS Spam"
[25] S.Gadde, A.Lakshmanrao,S.Satyanarayana, "SMS Spam Detection Using Machine Learning and Deep Learning Techniques", Computer Science ,2021 7th International Conference on Advance Computing and Communication Systems
[26] L .Zhang , J. Zhu, and T. Yao, "An evaluation of statistical spam filtering tecniques," ACM Transactions on Asian Language Information Processing (TALIP)
[27] M. Bassiouni, M. Ali, and E. A. El-Dashshan ,"Ham and spam E-mails classification using machine learning techniques," Journal of Applied Security Research
[28] A.Sharaff, "Comparative study of classification algorithms for spam email detection," in Emerging Research in Computing, Information, Communication and Applications
[29] S.Y. Bhatt," Spammer classification using ensemble methods over structural social network features," in Proceedings of the International Joint Conferences on Web Intelligence and Intelligent Agent Technologies.
[30] J.C Gomez and M.F-Moens," PCA document reconstruction for email classification," Computational Statistics and Data Analysis .



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY