International Journal of Engineering and Applied Physics (IJEAP)
Vol. 1, No. 3, September 2021, pp. 278~284
ISSN: 2737-8071

Machine learning-based Naive Bayes approach for divulgence of Spam Comment in Youtube station

Sohom Bhattacharya[1], Shubham Bhattacharjee[1], Anup Das[2], Anirban Mitra[3], Ishita Bhattacharya[4] , Subir Gupta[1]

[1]Department of Masters of Computer Application, Dr. B. C. Roy Engineering College, Durgapur, West Bengal – 713206, India
[2]CEO and founder of AnupTechTips
[3]Department of Computer Science & Engineering, ASETK, Amity University Kolkata, 700135, India.
[4]Department of Life Science, Binod Bihari Mahto Koyalanchal University, Dhanbad, 828130, India

Article Info

Article history:

Received May 10, 2021
Revised June 23, 2021
Accepted July 2, 2021

Keywords:

Artificial Intelligence
Naïve Bayes
Spam comment
Ham comment
QQ-plot

ABSTRACT

In the 21st Century, web-based media assumes an indispensable part in the interaction and communication of civilization. As an illustration of web-based media viz. YouTube, Facebook, Twitter, etc., can increase the social regard of a person just as a gathering. Yet, every innovation has its pros as well as cons. In some YouTube channels, a machine-made spam remark is produced on that recordings, moreover, a few phony clients additionally remark a spam comment which creates an adverse effect on that YouTube channel. The spam remarks can be distinguished by using AI (artificial intelligence) which is based on different Algorithms namely Naive Bayes, SVM, Random Forest, ANN, etc. The present investigation is focussed on a machine learning-based Naive Bayes classifier ordered methodology for the identification of spam remarks on YouTube

Corresponding Author:

Dr Subir Gupta
Assistant Prof MCA department
Dr. B.C Roy Engineering College
Durgapur, West Bengal –713206, India
subir2276@gmail.com

1. INTRODUCTION

Spam is the messages with inappropriate content i.e. it can be about adult videos, company sponsors, website, scam, invitations to different undesirable links, etc[1][2][3]. Moreover, these messages appear repeatedly on the comment page. In the YouTube videos, in the comments section, various types of spam or ham comments do appear. If it is not detected or filtered properly the particular channel probably loses its popularity as views, likes, dislikes, or followers of that video gets decreasing due to some annoying spam contents[4]. Most importantly, the YouTuber receives a handful of remuneration from YouTube for making more views and subscribers, but due to spam comments, the aforesaid remuneration sometimes gets hampered or lessened. Then again, it is capricious that the amount of the YouTube connections of any video are getting influenced because of spam content. Several research investigations indicating different spam detection techniques have been reported to detect spam with various methods[5][6][7][8]. A method for spam detection which can clarify fruitless and redundant features in blogs [9]. A runtime spam detection method known as BARS (Blacklist-Assisted Runtime Spam Detection) created a database of URLs containing spam against new post's URLs which can regulate the spam content in a post[10]. According to an investigation, a direct mapreduce algorithm to successfully locate spam masses[11]. Three different techniques to detect deceptive opinion spam: standard text classification, psycholinguistic deception detection [12]. In the present time, machine learning is one of the best research domains, it can be used in multi-disciplinary cases also like iris detection, traffic control, metallurgy, biomedical engineering and so many[13][14][15][16][17] .

2. BACKGROUND KNOWLEDGE AND DATASET PREPARATION

Different machine learning techniques are used for spam comment detection in YouTube videos viz. Random Forest, SVM, Logistics regression, Naive Bayes etc[18][19][20]. The naive Bayes approaches of the machine learning domain are most popularly used. The naive Bayes algorithm was invented by Thomas Bayes (1701- 1761). Naive Bayes Classifier is a collection of classification algorithm based on the Naive Bayes theorem. Naive Bayes is not a single algorithm but is a family of algorithms. The algorithms usually share a common rule that every pair of features being classified is autonomous of each other[21][22].

Naïve Bayes algorithms follows the

$P(Y / X) \alpha P(Y) * \prod_{i=1}^{n} \boldsymbol{P}(\boldsymbol{X}_{\boldsymbol{i}} / Y)$

Where the target of Naive Bayes was to discover the worth of Y which gives maximum probability.

The maximum value of Y will be Ymax = arg maxy [$P(Y) * \prod_{i=1}^{n} \boldsymbol{P}(\boldsymbol{X}_{\boldsymbol{i}} / Y)$]

Where ‘Arg max’ is an operation that finds the max value of Y

The dataset used for this investigation has been gotten from the general populace dataset store GitHub with URL:https://github.com/mohitgupta-omg/Kaggle-SMS Spam-Collection-Dataset-/mass/expert/spam.csv .In the given dataset, few comments have been noted as CSV record which was assembled from discrete worldwide notable and source-based YouTube channel. We have considered around 1200 comments.

3. METHODOLOGY

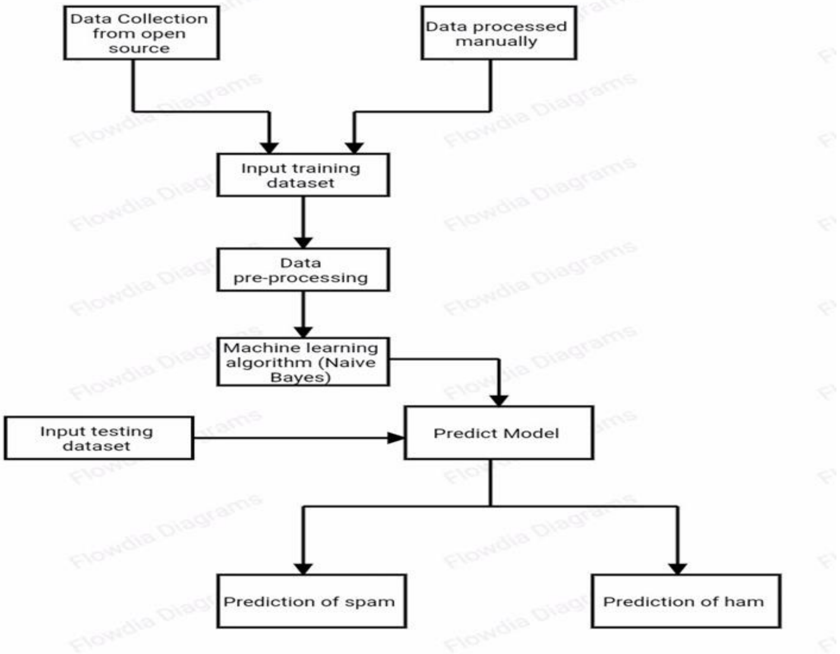

Figure 1. Flow Diagram of Spam Comment Detection using Naive Bayes Algorithm

The interaction utilized in Figure 1 for spam remark discovery has been expounded in this part. The previously mentioned informational collection establishes both spam and ham remarks. The all-out dataset has been combined and embedded for preparation. The preparation dataset has been pre-handled in the following stage and the equivalent has been handled through the Naïve Bayes calculation. A model has been anticipated and the testing information is embedded in the anticipated model to check the forecast aftereffect of the model. In this model the information dataset is ordered into two sections, first and foremost, utilized for preparing and also, for testing. The dataset utilized for preparing has been expounded in the next three stages.

In the first phase, the unorganized data gets converted into a more organized form which in term helps to filter the data.

The subsequent stage runs after the highlight assortment. If the said information is moved to Naive Bayes calculation, a few copy information can be acquired which is then changed over or limited to a bunch of capacity known as vector.

In the third stage, Count-Vectorizer is utilized to make a lattice where each particular word is addressed by the segment of the framework and the cell worth of the network addresses the coordinating with word check [23][24].

After the utilization of Count Vectorizer, information fitting has been refined. the information fitting ought to be executed so that it fabricates a model with the innocent Bayes calculation [25][26]. The information which is utilized as Counter-Vectorizer is being carried out by a capacity known as Multinomial NB which is reasonable for arrangement through Naive Bayes classifier and has been utilized in this exploration to acquire the normal outcome.

4. OBSERVATION AND RESULT

Table 1. Showing sample predict the result of spam comment

PREDICTIONS SPAM	ORIGINALS SPAM	COMMENTS
Spam	Spam	http://glearn.io/2z8qp.com
Ham	Ham	hey there you are
Spam	Ham	you are a nice guy
Ham	Ham	ðŸ‘•
Spam	Spam	get some nuts baby
Ham	Ham	look son there who is standing
Spam	Spam	I love you.com

Table 1, showing test foresee the consequence of spam remark of YouTube channel utilizing Naive Bayes classifier. Rejecting has been performed with next to no distinction in both as far as the precision level. Concerning the above table, one can say that a spam sifting strategy utilizing the Naive Bayes classifier for recognizable proof of spam comment on YouTube. gives a good and adequate outcome.

The predicted dataset by machine and manually scrapping has been performed with very little difference in both in terms of the accuracy level. With the assistance of this dataset, Bar diagram, Line Graph and QQ plot are made for accuracy checking displayed in underneath: -

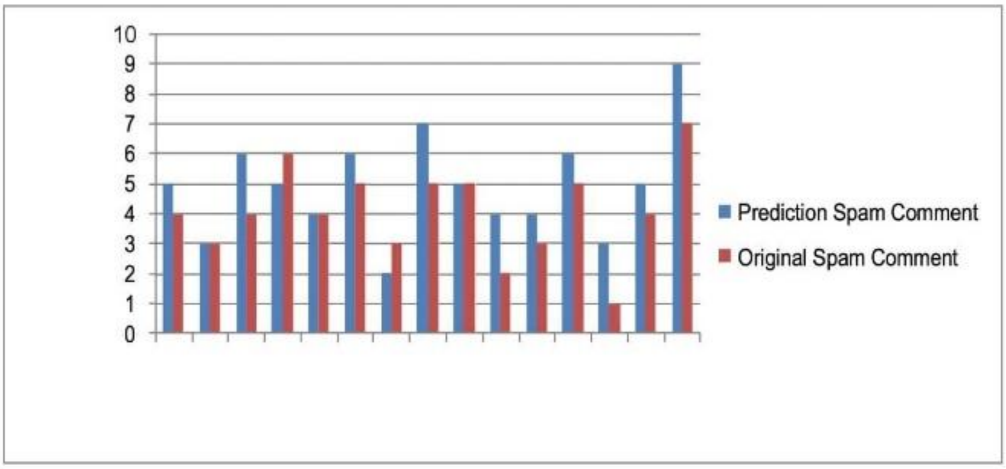

Figure 2. Bar graph representation of spam comment.

In the Figure 2 shown above represent a bar graph where x axis represents number of YouTube video result and y axis represent the no of spam comment, and the blue line shows the prediction spam comment and red line show the original spam comment and the graph is shown that we have gathered comment from 15 different YouTube channel and the data has been plotted which is shown that no of prediction comment and no of original comment.

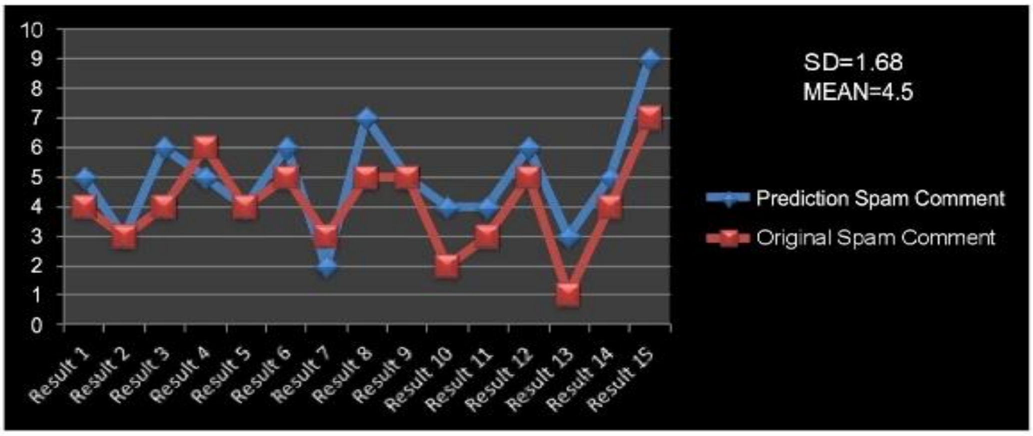

Figure 3. Line graph representation of spam comment.

In Figure 3: the x axis represent number of YouTube videos and y axis represents the number of spam comments. In this bar graph the blue line shows the number of prediction spam comment and the red line represents the original spam comments. Mean and SD has been calculated 4.5 and SD respectively.

Standard Deviation is defined as how the calculations for a group are scattered out from the average (mean or expected value). A low standard deviation implies that the vast majority of the numbers are near the average, while a high standard deviation implies that the numbers are more fanned out. Since the SD of this data is 1.68[38]. So, it can be concluded that the model has produced good result and the accuracy level of the model is approximately 98%.

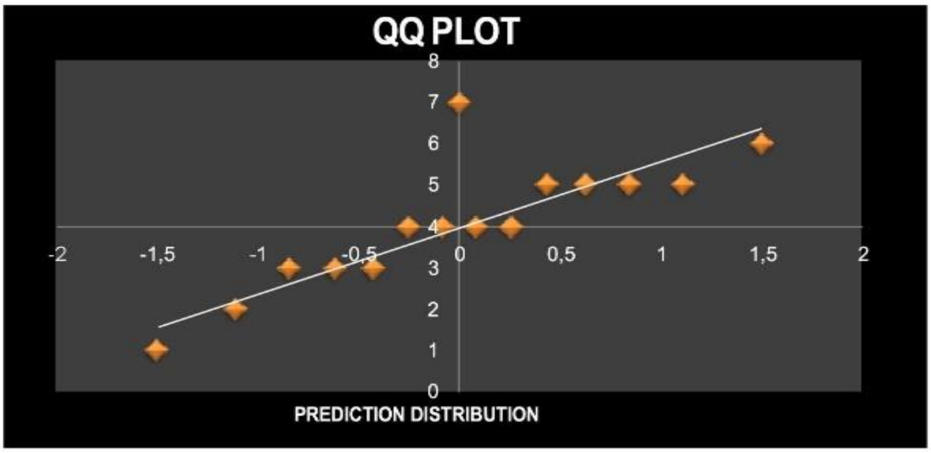

Figure 4. Scattered Graph Representation of Spam Comment using QQ Plot

In Figure 4 the x axis represents the prediction distribution and y axis represents the original distribution. The middle part of the chart forms a linear plot which means that the middle range of the prediction distribution Correctly map the middle range of the original distribution. The two ends of the graph represent the data in the tail is not linear

Table 2. Representation of Mean Square Error of spam detection of predict data and Original data.

	Prediction Spam Comment	Original Spam Comment	Error	Square Error
Result 1	5	4	-1	1
Result 2	3	3	0	0
Result 3	6	4	-2	4
Result 4	5	6	1	1
Result 5	4	4	0	0
Result 6	6	5	-1	1
Result 7	2	3	1	1
Result 8	7	5	-2	4
Result 9	5	5	0	0
Result 10	4	2	-2	4
Result 11	4	3	-1	1
Result 12	6	5	-1	1
Result 13	3	1	-2	4
Result 14	5	4	-1	1
Result 15	9	7	-2	4
				MSE = 1.8

Mean Square Error (MSE) is characterised as mean or average square of the distinction among genuine or accessed values. Value of MSE is always non- negative and if it is closer to zero then it is predicted as a good result but not in every case.

In table 2 the mean square error value of the total dataset is 1.8, so it can be concluded that the error value is very low. Due to which the error percentage of the predicted model is low and chances of error are very less. Therefore, the accuracy level of the model is very satisfying.

Table 3. Representation of t-test of spam detection of predict data and actual data

	Variable 1	Variable 2
Mean	4.933333	4.066667
Variance	3.066667	2.352381
Observation	15	15
Pearson Correlation	0.799595	
Hypothesized Mean		
Difference	0	
df	14	
t Stat	3.166295	
P(T<=t) one-tail	0.003433	
t Critical one-tail	1.76131	
P(T<=t) two-tail	0.006866	
t Critical two-tail	2.144787	

According to the dataset, t value and standard deviation have been calculated as 2.917035 and 1.68 respectively. The degree of freedom has been calculated as (number of observation-1). Since the p – value () is less than the alpha value (0.05), thus, the null hypothesis is rejected and it is proved that there is no significant difference in the means of each sample. It can be concluded that the prediction is near to accuracy.

5. CONCLUSION

The issue related to spam is growing significantly all through the world. With misrepresented globalization and the right to speak freely of discourse, the abuses of the equivalent are very obvious inside the sort of spam remarks on changed computerized stages. The current exploration examination portrays a spam sifting technique utilizing the Naive Bayes classifier, intending to improve the idea of substance in the web.

REFERENCES

[1] S. Aiyar and N. P. Shetty, “N-Gram Assisted Youtube Spam Comment Detection,” Procedia Comput. Sci., vol. 132, no. Iccids, pp. 174–182, 2018, doi: 10.1016/j.procs.2018.05.181.

[2] C. A. Shue, M. Gupta, C. H. Kong, J. T. Lubia, and A. S. Yuksel, “Spamology: A study of spam origins,” 6th Conf. Email Anti-Spam, CEAS 2009, no. January, 2009.

[3] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, “Detection of review spam: A survey,” Expert Syst. Appl., vol. 42, no. 7, pp. 3634–3642, 2015, doi: 10.1016/j.eswa.2014.12.029.

[4] 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,” J. Netw. Comput. Appl., vol. 112, pp. 53–88, 2018, doi: 10.1016/j.jnca.2018.03.015.

[5] Z. Guo, L. Tang, T. Guo, K. Yu, M. Alazab, and A. Shalaginov, “Deep Graph neural network-based spammer detection under the perspective of heterogeneous cyberspace,” Futur. Gener. Comput. Syst., vol. 117, pp. 205–218, 2021, doi: 10.1016/j.future.2020.11.028.

[6] D. C. Corrales, A. Ledezma, and J. C. Corrales, “A case-based reasoning system for recommendation of data cleaning algorithms in classification and regression tasks,” Appl. Soft Comput. J., vol. 90, p. 106180, 2020, doi: 10.1016/j.asoc.2020.106180.

[7] A. Fahfouh, J. Riffi, M. Adnane Mahraz, A. Yahyaouy, and H. Tairi, “PV-DAE: A hybrid model for deceptive opinion spam based on neural network architectures,” Expert Syst. Appl., vol. 157, p. 113517, 2020, doi: 10.1016/j.eswa.2020.113517.

[8] S. Panda, A. K. Ghosh, A. Das, U. Dey, and S. Gupta, “Machine Learning-based Linear regression way to deal with making data science model for checking the sufficiency of night curfew in Maharashtra , India,” vol. 1, no. 2, pp. 168–173, 2021.

[9] A. Kantchelian, J. Ma, and A. D. Joseph, “Robust Detection of Comment Spam Using Entropy Rate Categories and Subject Descriptors,” no. AISec, pp. 59–69, 2012.

[10] E. Tan, L. Guo, S. Chen, X. Zhang, and Y. E. Zhao, “Spammer Behavior Analysis and Detection in User Generated Content on Social Networks,” 2012, doi: 10.1109/ICDCS.2012.40.

[11] Advances in Intelligent Systems. .

[12] J. T. Hancock and C. Cardie, “Finding Deceptive Opinion Spam by Any Stretch of the Imagination Finding Deceptive Opinion Spam by Any Stretch of the Imagination,” no. May, 2014.

[13] S. Adamović et al., “An efficient novel approach for iris recognition based on stylometric features and machine learning techniques,” Futur. Gener. Comput. Syst., vol. 107, pp. 144–157, 2020, doi: 10.1016/j.future.2020.01.056.

[14] S. Lee et al., Intelligent traffic control for autonomous vehicle systems based on machine learning, vol. 144. Elsevier Ltd, 2020.

[15] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, “Machine learning for predictive modelling based on small data in biomedical engineering,” IFAC-PapersOnLine, vol. 28, no. 20, pp. 469–474, 2015, doi: 10.1016/j.ifacol.2015.10.185.

[16] S. Gupta et al., “Modelling the steel microstructure knowledge for in-silico recognition of phases using machine learning,” Mater. Chem. Phys., vol. 252, no. March, p. 123286, 2020, doi: 10.1016/j.matchemphys.2020.123286.

[17] S. Gupta, J. Sarkar, M. Kundu, N. R. Bandyopadhyay, and S. Ganguly, “Automatic recognition of SEM microstructure and phases of steel using LBP and random decision forest operator,” Meas. J. Int. Meas. Confed., vol. 151, p. 107224, 2020, doi: 10.1016/j.measurement.2019.107224.

[18] N. N. Amir Sjarif, N. F. Mohd Azmi, S. Chuprat, H. M. Sarkan, Y. Yahya, and S. M. Sam, “SMS spam message detection using term frequency-inverse document frequency and random forest algorithm,” Procedia Comput. Sci., vol. 161, pp. 509–515, 2019, doi: 10.1016/j.procs.2019.11.150.

[19] Y. Tian, M. Mirzabagheri, P. Tirandazi, and S. M. H. Bamakan, “A non-convex semi-supervised approach to opinion spam detection by ramp-one class SVM,” Inf. Process. Manag., vol. 57, no. 6, p. 102381, 2020, doi: 10.1016/j.ipm.2020.102381.

[20] B. K. Dedeturk and B. Akay, “Spam filtering using a logistic regression model trained by an artificial bee colony algorithm,” Appl. Soft Comput. J., vol. 91, p. 106229, 2020, doi: 10.1016/j.asoc.2020.106229.

[21] N. M. Samsudin, C. F. B. Mohd Foozy, N. Alias, P. Shamala, N. F. Othman, and W. I. S. Wan Din, “Youtube spam detection framework using naïve bayes and logistic regression,” Indones. J. Electr. Eng. Comput. Sci., vol. 14, no. 3, pp. 1508–1517, 2019, doi: 10.11591/ijeecs.v14.i3.pp1508-1517.

[22] C. C. Kiliroor and C. Valliyammai, Social context based naive bayes filtering of spam messages from online social networks, vol. 758. Springer Singapore, 2018.

[23] C. M. Yeomans, R. K. Shail, S. Grebby, V. Nykänen, M. Middleton, and P. A. J. Lusty, “A machine learning approach to tungsten prospectivity modelling using knowledge-driven feature extraction and model confidence,” Geosci. Front., vol. 11, no. 6, pp. 2067–2081, 2020, doi: 10.1016/j.gsf.2020.05.016.

[24] V. Zorkadis, D. A. Karras, and M. Panayotou, “Efficient information theoretic strategies for classifier combination, feature extraction and performance evaluation in improving false positives and false negatives for spam e-mail filtering,” Neural Networks, vol. 18, no. 5–6, pp. 799–807, 2005, doi: 10.1016/j.neunet.2005.06.045.

[25] L. Yang et al., “Prediction model of the response to neoadjuvant chemotherapy in breast cancers by a Naive Bayes algorithm,” Comput. Methods Programs Biomed., vol. 192, 2020, doi: 10.1016/j.cmpb.2020.105458.

[26] J. Kolluri and S. Razia, “Text classification using Naïve Bayes classifier,” Mater. Today Proc., no. xxxx, 2020, doi: 10.1016/j.matpr.2020.10.058.

BIOGRAPHIES OF AUTHORS

	Sohom Bhattacharya (Completed M.C.A. Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India)
	Shubham Bhattacharjee (Completed M.C.A. Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India)
	Anup Das (CEO and founder of AnupTechTips)
	Anirban Mitra (Professor of Computer Science & Engineering department Amity University Kolkata, 700135, India.)
	Dr Ishita Bhattacharya (Professor of department of Life Science Binod Bihari Mahto Koyalanchal University, Dhanbad, 828130, India)
	Dr Subir Gupta (Assistant Prof MCA department Dr. B.C Roy Engineering College Durgapur, West Bengal –713206, India)