Journal of Soft Computing Exploration

Performance comparison of support vector machine and gaussian naive bayes classifier for youtube spam comment detection

Yahya Nur Ifriza[1], Muhammad Sam'an[2]
[1]Department of Computer Science, Universitas Negeri Semarang, Indonesia
[2]Postgraduate Student, Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Malaysia

Article Info

Article history:

Received May 20, 2021
Revised Aug 19, 2021
Accepted Aug 29, 2021

Keywords:

Youtube spam
Text mining
Gaussian naïve bayes
Support vector machine
Machine learning

ABSTRACT

Youtube is a video-sharing website that was launched in 2005 and has been around ever since. Youtube produces over 400 hours of substance each moment and more than 1 billion hours of substance are devoured by clients every day. In this work, we present a new approach by comparing the analysis results using a support vector machine and the Gaussian Naive Bayes classificatio. Our proposed methodology We used the dataset from UCI especially Youtube-Shakira for testing and training purposes. In Naive Bayes and SVM, the altered dataset is separated into training and testing subsets and supplied to them. In all cases, the F1 score was used to evaluate the classifier's performance. The results of the experiment are displayed in Gaussian Naive Bayes with an F1 score of 84.38% and a Support Vector Machine (SVM) with an F1 score of 88.00%. Naive Bayes is consistently the worst performer than SVM.

Corresponding Author:

Yahya Nur Ifriza,
Department of Computer Science,
Universitas Negeri Semarang,
Sekaran, Gunungpati, Semarang 50229, Indonesia.
Email: yahyanurifriza@mail.unnes.ac.id

1. INTRODUCTION

Youtube is a type of social media on the Web for video sharing which was launched in 2005, then acquired by Google in 2006 and until now has become one of Google's subsidiaries which is growing very rapidly in the world. YouTube emerged as the highest rated competitor for video sharing platforms. A YouTuber is a person who uploads videos to YouTube. They can rate, comment and subscribe to other YouTubers, as well as rate and share their own videos. More than 2 billion logged-in people accessed YouTube every month, and 1 billion hours of video material were watched every day, leading to billions of views by 2020, according to Youtube [1][2].

One of the most widely used features of the YouTube platform is the YouTube comment column, where users can give reactions in the form of suggestions, criticism or simply express their liking for videos uploaded on this platform. It is possible, however, that this function may be exploited to spread hate speech and criminal activity, or to disseminate commercial information, commonly known as spam. Many of the spam comments on YouTube have nothing to do with the substance of the video and are generally produced by automated bots. O'Callaghan et al. [3] said that the ability of robots to campaign massively hate speech or crime on a large scale and well-organized. As reported by Cassin et al. [4] in BBC that YouTube has come

under serious criticism for its inability to classify uploadedcontent based In fact, a large part of Youtube's user base comes from a young age so children are vulnerable to being exposed to harmful material through spam comments.

Aiyar and Shetty [5] classified YouTube spam comment based on its origin, i,e. (1) Link based is contained Hypertext links to other web-sites that are often found on YouTube content itself and many links often direct users to dangerous sites without user notifications; (2) Channels promotion is a type of spam comment that users create by promoting their own channel to ask customers, to post a link to their video. Example of spam comments based on type are shown in Table 1. YouTube has also tried to block spam in the comments column in the form of a link containing Hypertext links. Even though this method is considered effective. However, fake account users or spammers be more creative again. They are looking for another way to insert a whitespace character between the links to avoid detection.

Table 1. Spam comment based on type

Nature	Example
Link Based	make your iPhone 6 / 6s happy http://www.ebay.com/itm/272739565815?ulnoapp=true I HAVE SOME THING OWSOME FOR YOU, I'M SURE YOU LIKE IT, IT'S OWSOME
Channel Promotion	Pls some one help me with my channel like can anyone just help by giving me a few subs am trying to come back Go checkout my channel and subscribe I will subscribe back

Many researchers had been studied the classification of YouTube comments as spam or ham by using machine learning. Kantchelian et al. [6] developed a logistic regression based entropy rate. Aziz et al. [7] discussed the performance comparison of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Alias et al. [8] used six classifier of machine learning techniques i.e Random Tree (RT), Random Forest (RF), Naive Bayes, KStar, Decision Table and Decision Stump for YouTube live streaming spam comments detection. In this work, we present the performance comparison of support vector machine and the Gaussian Naive Bayes classification for Youtube spam comment detection. However, we ended up achieving comparable results and the best accuracy in spam data filtering practices. This may be managed by analyzing the underlying data sets allowing users to distinguish spam data that is displayed as true or false.

2. RELATED WORK

Aiyar dan Shetty [5] video-sharing website, Youtube, to detect unwanted remarks or spam. For example, they proposed expanding their effort to include URLs and short message elimination as well as employing N-grams which have been shown to be highly efficient in detecting and then combatting spam comments.

Al-Zoubi et al. [9] their algorithm to detect spam profiles is considered as one of the most challenging issues in online social networks, the experiments and results show that the proposed model outperforms many other algorithms in terms of accuracy, and provides very challenging results in terms of precision, recall, f-measure and AUC.

Boyd [10] A revision of the conventional participatory framework categories is offered on the basis of the new online settings. A multilevel depiction of production is proposed, with Obama's speech as the first level of production and his remarks as the second level of production.

Chakraborty et al. [11] Reviewed current advancements in social spam detection and mitigation strategies, its theoretical models and implementations, along with their qualitative comparison.

Makkar and Kumar [12] in IoT environment, devised an effective deep learning-based technique for online spam detection WebSPAMUK 2007 was used to validate the proposed system. For pre-processing, the dataset is split by over-sampling and then trained using a novel approach called "Underfitting".

Sing et al. [13] when it comes to the identification of false news, they find that Bernoulli's Naive Bayes Classifier outperforms Gaussian Naive Bayes in terms of classification results.

Yin et al. [14] Our findings on a real-world multi-relational social network indicate the efficiency of our proposed MDM on multi-relational social spammer detection by utilizing multi-level dependence of relational sequences.

3. METHOD

These sections include a breakdown of the data collected and compiled, as well as a description of the cycle used to rank the comments as Spam. Figure 1 depicts a flowchart of the complete procedure in simple terms. As a training and testing dataset, we used the UCI machine learning repository, namely Youtube-Shakira. In Naive Bayes and SVM, the modified dataset is separated into training and testing subsets and given to them. [15]–[17]. Python 3.9.2 and the NLP library were used for all implementations. In this methodical interaction, a three-stage rule, to be specific planning, conducting, and documenting [18][19][20][21]. The three phases of the research methodology are discussed as follows.

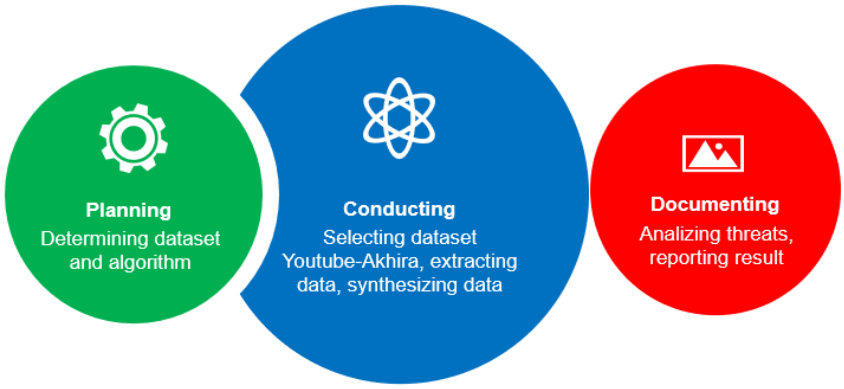

Figure 1. Research methodology

Figure 1 shown the complete procedure in a few words, on planning process we determining dataset and algorithm to get the best plan and than on conducting we must selecting dataset Youtube-Shakira, extracting data and synthezing data to became data train and data test, on the last documenting process to analizing threats and reporting result. On the figure 2 shows flowchart taxonomy of big data analysis.

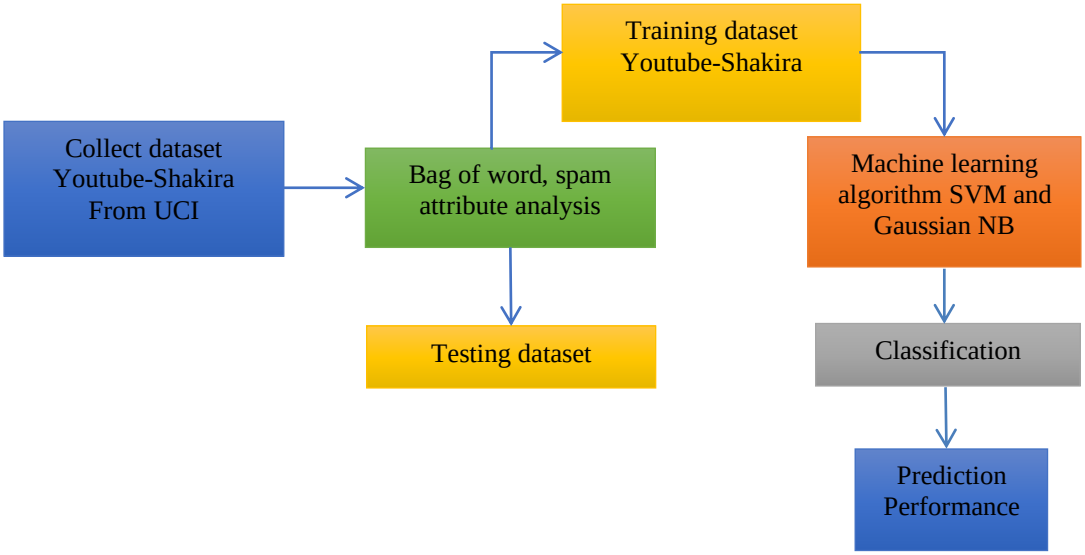

Figure 2. Flowchart taxonomy of big data nalysis

We removed around 20.000 remarks from Youtube-Shakira on UCI machine learning repository and put away them in an information base for additional examination. Specifically, moving music recordings with extremely huge perspectives were focused on the grounds that we realize that Youtube-Shakira have 32.500.000 subscriber. Straightforward library channel is utilized to identify and extricate remarks with Latin letters just as its motivation is to assess English remarks. Because of the generally low proportion of spam remarks, an essential hand-designed spam channel is utilized to extricate possibly malicious remarks. The handmade spam channel comprises of a progression of basic customary articulations containing a spam-based remark component. So our last dataset contains a generally equivalent proportion for spam. We name remarks that are limited time or wrong with explicit recordings and order them as spam.

Multinomial Naive Bayes (M-NB) and Support Vector Machine (SVM) were used independently for preparation and classification. Naive Bayes was chosen as the standard because of its simplicity and great efficacy. [18]. Support Diverse research have shown that vector machines are excellent for characterisation difficulties. [22]–[24]. As a result of Natural Language Processing, SVM's unique approach is ideally suited for large datasets with a lot of dimensions.

4. RESULTS AND DISCUSSIONS

Using the cross-approval and k-fold method, we evaluate the display of our spam remark discovery framework. A random number is used to reorder the data. Cross-approval was performed using a five-crease crossapproval method. The whole dataset was divided into five equal sections, with one part being used as the test set and the rest as the preparation set in each overlay. When the effects of each overlap are averaged together, we get the F1 Score for the last time. As part of the computation, the classifier's precision is compared with the testing subset.

Matthews correlation coefficient and F1 Score were employed as measures for assessment. The algorithm's performance, however, is evaluated using the F1 score since we want to achieve both high accuracy and high specificity.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Specificity = \frac{FN}{TN+FN} \quad (2)$$

$$F1\ score = 2 . \frac{Precision.recall}{precision+recall} \quad (3)$$

where TP, TN, FP, FN in Eq (1), (2) and (3) represent the true positive, false positive, and false negative rates.

Using backend sequential backend with 1 concurrent worker, install 10 folds with 1 candidate each. A classifier's performance was always assessed using the F1 score. 84.38 percent in Gaussian Naive Bayes and 88.00 percent in Support Vector Machine (SVM) are the outcomes of the experiment. It is evident from the table data that Naive Bayes regularly outperforms the more sophisticated method of SVM.

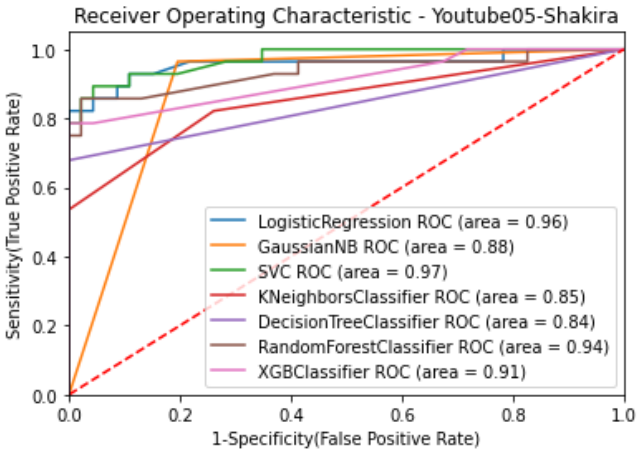

Figure 3. Receiver operating characteristic

Figure 3 the Receiver operating characteristic of sensitivity score (Y-axis) vs the specificity test results (X-axis). The area Gaussian Naive Bayes score of 0.88 and the Support Vector Machine score of 0.97 have both increased considerably in the last year or more.

He examined current advancements in social spam detection and mitigation approaches, its theoretical models and implementations, as well as a qualitative comparison between them, according to Chakraborty [11], but they don’t give comparison some algorithm classification to know accuracy performance. The performance F1 scores for the Gaussian Naive Bayes and Support Vector Machine values can be seen in Table 1 below.

Table 1. F1 scores for GaussianNB and SVM

Classifier	Accuracy	Spam Caught	Blocked Ham	Mathews Coeff	F1 score
GaussianNB	86.49	75.00	2.63	74.58	84.38
SVM	91.89	100.00	11.54	83.37	88.00

5. CONCLUSION

In this work, we offer a technique for automatic machine-assisted spam comment identification on the Youtube-Shakira platform, and we demonstrate the efficiency of utilizing GaussianNB and SVM on performance classifications. We see that SVM with F1 score 88.00 is better than GaussianNB with F1 score 84.38 at identifying the size of spam in comments. As confirmed [5], Support to classify large data sets, Vector Machines and NGram outperform other standard Machine Learning methods. A better categorization representation will be possible with the development of false news classification implementation in future work.

REFERENCES

[1] YouTube for Press, "YouTube for Press," *YouTube*, 2017. .

[2] K. Budiman, A. T. Putra, Alamsyah, E. Sugiharti, M. A. Muslim, and R. Arifudin, "Implementation of ERP system functionalities for data acquisition based on API at the study program of Universities," *J. Phys. Conf. Ser.*, vol. 1918, no. 4, 2021, doi: 10.1088/1742-6596/1918/4/042151.

[3] D. O'Callaghan, M. Harrigan, J. Carthy, and P. Cunningham, "Network analysis of recurring YouTube spam campaigns," in *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, 2012, pp. 531–534.

[4] E. Cassin, A. Subedar, and M. Wendling, "Glitch in YouTube's tool for tracking obscene comments," 2017. .

[5] 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.

[6] A. Kantchelian, J. Ma, L. Huang, S. Afroz, A. D. Joseph, and J. D. Tygar, "Robust detection of comment spam using entropy rate," in *Proceedings of the ACM Conference on Computer and Communications Security*, 2012, pp. 59–69, doi: 10.1145/2381896.2381907.

[7] A. Aziz, C. F. Mohd Foozy, P. Shamala, and Z. Suradi, "YouTube Spam Comment Detection Using Support Vector Machine and K–Nearest Neighbor," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 2, p. 612, 2018, doi: 10.11591/ijeecs.v12.i2.pp612-619.

[8] N. Alias, C. F. M. Foozy, and S. N. Ramli, "Video spam comment features selection using machine learning techniques," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 15, no. 2, pp. 1046–1053, 2019, doi: 10.11591/ijeecs.v15.i2.pp1046-1053.

[9] A. M. Al-Zoubi, H. Faris, J. Alqatawna, and M. A. Hassonah, "Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts," *Knowledge-Based Syst.*, vol. 153, pp. 91–104, 2018, doi: 10.1016/j.knosys.2018.04.025.

[10] M. S. Boyd, "(New) participatory framework on YouTube? Commenter interaction in US political speeches," *J. Pragmat.*, vol. 72, pp. 46–58, 2014, doi: 10.1016/j.pragma.2014.03.002.

[11] M. Chakraborty, S. Pal, R. Pramanik, and C. Ravindranath Chowdary, "Recent developments in social spam detection and combating techniques: A survey," *Inf. Process. Manag.*, vol. 52, no. 6, pp. 1053–1073, 2016, doi: 10.1016/j.ipm.2016.04.009.

[12] A. Makkar and N. Kumar, "An efficient deep learning-based scheme for web spam detection in IoT environment," *Futur. Gener. Comput. Syst.*, vol. 108, pp. 467–487, 2020, doi: 10.1016/j.future.2020.03.004.

[13] M. Singh, M. Wasim Bhatt, H. S. Bedi, and U. Mishra, "Performance of bernoulli's naive bayes classifier in the detection of fake news," *Mater. Today Proc.*, no. xxxx, 2020, doi: 10.1016/j.matpr.2020.10.896.

[14] J. Yin, Q. Li, S. Liu, Z. Wu, and G. Xua, "Leveraging multi-level dependency of relational sequences for social spammer detection," *Neurocomputing*, vol. 428, pp. 130–141, 2020.

[15] M. Ontivero-Ortega, A. Lage-Castellanos, G. Valente, R. Goebel, and M. Valdes-Sosa, "Fast Gaussian Naïve Bayes for searchlight classification analysis," *Neuroimage*, vol. 163, pp. 471–479, 2017, doi: 10.1016/j.neuroimage.2017.09.001.

[16] A. Khajenezhad, M. A. Bashiri, and H. Beigy, "A distributed density estimation algorithm and its application to naive Bayes classification," *Appl. Soft Comput.*, vol. 98, p. 106837, 2021, doi: 10.1016/j.asoc.2020.106837.

[17] Q. He *et al.*, "Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms," *Sci. Total Environ.*, vol. 663, pp. 1–15, 2019, doi: 10.1016/j.scitotenv.2019.01.329.

[18] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, 2007, doi: 10.1016/j.jss.2006.07.009.

[19] M. Sam and Y. N. Ifriza, "A combination of TDM and KSAM to determine initial feasible solution of transportation problems," *J. Soft Comput. Explor.*, vol. 2, no. 1, pp. 17–24, 2021, doi: 10.52465/joscex.v2i1.16.

[20] Y. N. Ifriza and M. Sam, "Irrigation management of agricultural reservoir with correlation feature selection based binary particle swarm optimization," *J. Soft Comput. Explor.*, vol. 2, no. 1, pp. 40–45, 2021, doi: 10.52465/joscex.v2i1.23.

[21] Y. N. Ifriza, C. E. Edi, and J. E. Suseno, "Expert system irrigation management of agricultural reservoir system using analytical hierarchy process (AHP) and forward chaining method," *Proc. of ICMSE*, pp. 74–83, 2017.

[22] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *J. Mach. Learn. Res.*, vol. 15, no. October, pp. 3133–3181, 2014, doi: 10.1117/1.JRS.11.015020.

[23] I. Chaturvedi, E. Cambria, R. E. Welsch, and F. Herrera, "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," *Inf. Fusion*, vol. 44, no. December 2017, pp. 65–77, 2018, doi: 10.1016/j.inffus.2017.12.006.

[24] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, and E. Cambria, "Bayesian network based extreme learning machine for subjectivity detection," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1780–1797, 2018, doi: 10.1016/j.jfranklin.2017.06.007.