AN INNOVATIVE METHOD TO ENHANCE THE ACCURACY IN CLASSIFICATION OF SPAM DETECTION FOR YOUTUBE COMMENTS WITH USING SUPPORT VECTOR MACHINE OVER K–NEAREST NEIGHBOR

S. Venkata Manoj Kumar Reddy[1], K. Malathi [2]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The aim of this research article is to elaborate the accuracy rate for detecting the spam comments on YouTube videos by using Innovative Support Vector Machine (SVM) in comparison with K-Nearest Neighbor (KNN) Classifier.

Materials & Methods: The data set in this paper utilizes UCI machine learning repositories. The sample size of predicting the spam comments on YouTube videos with enhanced accuracy rate was sample 80 (Group 1=40 and Group 2 =40) and calculation is performed utilizing G-power 0.8 with alpha and beta qualities are 0.05, 0.2 with a confidence interval at 95%. Predicting the spam comments on YouTube videos with enhanced accuracy rate is performed by Innovative Support Vector Machine (SVM) whereas number of samples (N=10) and K-Nearest Neighbor (KNN) where number of samples (N=10).

Results and Discussion: The Innovative Support Vector Machine (SVM) classifier has 93.047 higher accuracy rates when compared to the accuracy rate of K-Nearest Neighbor (KNN) is 87.73. The study has a significance value of $p<0.05$ i.e. $p=0.022$.

Conclusion: Innovative Support Vector Machine (SVM) provides the better outcomes in accuracy rate when compared to K-Nearest Neighbor (KNN) for detecting the spam comments on YouTube videos with enhanced accuracy rate.

Keywords: Spam Detection, Innovative Support Vector Machine, K-Nearest Neighbor Classifier, Accuracy Rate, Youtube Spam, Machine Learning.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, pincode:602105.

[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, pincode:602105.

1. Introduction

The objective of the research paper is to elaborate the accuracy rate for detecting the spam comments on YouTube videos with improved accuracy rate. Online social networks such as Face-book and YouTube have become increasingly popular outlets over the past few years and are an integral part of the everyday life of people(Grabowicz et al. 2012). People spend a lot of time posting their messages on micro blogs websites, sharing their thoughts and making friends around the world. YouTube is ranked by teenagers as the most successful social network(Figueiredo, Benevenuto, and Almeida 2011). Currently 200 million users create up to four hundred million new YouTube videos every day. This rapid expansion of the YouTube platform encourages more spammers to create spam on the YouTube platform that contains malicious links to external websites that contain malware downloads, phishing, drug sales, or scams(Sharmin and Zaman 2017). One of YouTube's most well-known features is its comment system, in which different users may comment on uploaded videos. YouTube users have channels of their own. YouTube lets channels upload, share, rate, post comments on videos and subscribe to different channels. This remarkable feature allows users to communicate with each other and express their emotions, thoughts, etc. on the content. But, this has also made it possible for malicious users to post irrelevant or self-promotional content often called spam(Zhang et al. 2017). Spam comments are mostly entirely unrelated to the video and are usually created by automated bots posing as legitimate users. The main objective of this research work is to provide a prediction so that it can perform descriptive analytics on prediction of spam comments in YouTube. This study proposes the classification algorithm called support vector machine (SVM) for classifying spam comments. Experimental results show that the Innovative Support Vector Machine (SVM) shows good accuracy results.

Numerous researches have been done on predicting spam comments on YouTube videos using machine learning algorithms in recent years. IEEE explore published 65 research papers, and Google Scholar found 53 articles. In a study conducted by Yusof and Sadoon(Yusof and Sadoon 2017), a framework used consists of five (5) phases which is data collection, pre-processing, feature construction, spam detection and evaluation. The framework used by Alberto et al.,(Túlio C. Alberto, Lochter, and Almeida 2015) has three (3) phases such as processed data, pre-processing and classification. While the framework in research conducted by Kiran(Kiran 2015) has three (3) phases consisting of data collection, feature selection and classification. Besides that, the framework used in research by Chowdury et al.,(Chowdury et al. 2013) consists of data collection, select attribute and classifications. Saumya Goyal et al.(Goyal, Chauhan, and Parveen 2016) spam message detection on real twitter social media dataset is applied using KNN and decision tree. R.S.Yenape et al.(Kanodia, Sasheendran, and Pathari 2018) focuses on clearing out spam videos and at the same time also identifies the spammers responsible. It gathers the required attributes for the video in question with the help of an API which takes the URL of the video as input and crawls the page to obtain the necessary information. Finally, classification is done with ID3, K Nearest Neighbours and Naive Bayes algorithms. The existing study in YouTube Spam Detection is Manwar(Manwar, Lambhate, and Patil 2017) and Alberto(Tulio C. Alberto, Lochter, and Almeida 2015) show that both of the authors used Innovative Support Vector Machine (SVM) as a classifier in classification phase. Manwar(Manwar, Lambhate, and Patil 2017) stated that SVM classification is in binary-two class. Usually, class denoted by 0 and 1. However, the collection data have been classified into two classes. The most cited article was(Rădulescu, Dinsoreanu, and Potolea 2014) in IEEE Explore with 9 citations and 789 full text views.

Our team has extensive knowledge and research experience that has translated into high quality publications (K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022). The main problem of the existing method is that the accuracy depends on the quality of the data and the prediction stage is slow because of large amounts of information. As training data increases, the speed at which calculations are made rapidly decreases. This paper proposed a novel Innovative support vector machine (SVM) classifier to detect the spam comments on YouTube videos and compares the results with the K-Nearest Neighbor (KNN) classifier on the basis of accuracy, precision, and recall values. The main objective of this paper is to evaluate the accuracy of the SVM and K-NN classifier before and after optimizing the important

parameters using the Python software tool. The performance analysis of the proposed spam comment detection method gives better accuracy results than the existing K-NN method.

2. Methods and Materials

This work was carried out in the Digital Image Processing Laboratory, Department of Computer Science and Engineering, Saveetha School of Engineering. In this paper, The YouTube Spam Collection Data Set Collect from UCI machine learning repositories(Lichman and Others 2013). The dataset contained ten selected videos and were downloaded from YouTube through API. It is composed of 1,500 real and non-encoded messages that were labeled as legitimate (ham) or spam. Each sample represents a text comment posted in the comments section of each selected video. Group 1 was a K-Nearest Neighbor (KNN) algorithm and Group 2 was an innovative novel support vector machine (SVM) model. Python 3.9.2 and the NLP library were used for all implementations. The calculation is performed utilizing G-power 0.85 with alpha and beta qualities 0.05, 0.1 with a confidence interval at 95%.

K-Nearest Neighbor

K-NN is a supervised learning method. KNN is the simple and useful classifier for different classification problems. It does not require any prior knowledge of training like SVM or other machine learning algorithms. If the new training pattern is affixed to the subsisting training set then it doesn't require retraining. Before classifying a new element vector, a comparison should be finished with the training sample using distance metrics. Its k-nearest neighbors are then considered where the class that occurs most among the neighbors is given to the element that to be classified. A new element is classified on the basis that the neighbors are weighted by the distance measure. The appropriate working of the scheme depends on the proper selection of the appropriate parameter such as the 'k' which represents the number of neighbors used to give a class to any new element. The KNN only classifies the data into two classes which is very inefficient and reduces accuracy of classification. The KNN classifier is needed to improve spam comments identification accuracy through more advanced machine learning and deep learning algorithms. In pattern recognition, the K-Nearest Neighbors algorithm (or k-NN for short) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

- In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.
- In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors. The sample group 1 is the K-Nearest Neighbor (KNN) algorithm is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. The steps involved in the implementation of the KNN algorithm are described as follows.

Support Vector Machine

The SVM will refer to both classification and regression methods and the terms Support Vector Classification (SVC) and Innovative Support Vector Regression (SVR) will be used for specification. This part will continue with a complete introduction to structural risk(Ingale and Mahajan, n.d.). The SVM is introduced in the setting of classification, being both historical and more usable. This leads onto mapping the input into a higher dimensional feature space by a required choice of kernel function. The report then considers the problem of regression. SVM has successful applications in Bioinformatics, Text & Image Recognition. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one "target value" (i.e. the class labels) and several "attributes" (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. Support vectors are information focuses that are nearer to the hyperplane and impact the position and direction of the hyperplane. Utilizing these Innovative support vectors, we augment the edge of the classifier. Erasing the Innovative support vectors will change the situation of the hyperplane. These are the focuses that help us assemble our SVM. The solicitation approach is advantageous to pick at time spans somewhere around two dynamically potential results depending upon unsurprising pointing factors. In two-dimensional space this hyperplane is a line partitioning a plane in two sections where in each class lay in one or the other side. In multidimensional

space, the partition of the class is a hyperplane. AInnovative support vector machine is an uncommon preparation system that isolates measurements and withdraws both request and backslide methodology material plans. The request approach is worthwhile to pick at time periods at least two progressively potential outcomes relying upon predictable pointing factors(Gaye, Zhang, and Wulamu 2021). The SVM estimation gives out the objective information into any of these classes provided, considering information. The check of SVM is utilized on straightforward information based speed and not really determining. Determination tree models are memory-packed understudies in this manner they are not fitting for colossal sets, after which we work on cost-delicate home loan obligation assumption when using two-class support vector machine and switch over to resemble sets on the repeated enlightening machine a brief time frame later. SVM is successfully suitable in differentiating positive and negative problems such as spam.

The sample preparation group 2 is the novel Innovative support vector machine (SVM) algorithm, which is a supervised machine learning model that uses classification algorithms for two-group classification problems. It is used to estimate probability whether an instance belongs to a class or not. The experimental results show that the proposed SVM method has achieved better accuracy results.

Statistical Analysis

The output is obtained by using Python software. To train these datasets, required a monitor with resolution of 1024×768 pixels (7th gen, i5, 4 8GB RAM, 500 GB HDD), and Python software with required library functions and tool functions. For statistical implementation, the software tool used here is IBM SPSS V26.0(Hilbe 2004). The independent sample t test was performed to find the mean, standard deviation and the standard error mean statistical significance between the groups, and then comparison of the two groups with the SPSS software will give the accurate values for the two different s which will be utilized with the graph to calculate the significant value with maximum accuracy value (83.04%), mean value (83%) and standard deviation value (0.23323). Dependent variables are accuracy and independent variables are KNN and SVM methods.

3. Results

Figure.1 shows the simple bar graph for K-Nearest Neighbor (KNN) Classifier accuracy rate is compared with Support Vector Machine (SVM) Classifier. The Innovative Support Vector Machine (SVM) Classifier is higher in terms of accuracy rate 93.047 when compared with K-Nearest Neighbor (KNN) Classifier 87.73. Variable results with its standard deviation ranging from 80 lower to 90 higher K-Nearest Neighbor (KNN) Classifier where Support Vector Machine (SVM) Classifier standard deviation ranging from 90 lower to 100 higher. There is a significant difference between the K-Nearest Neighbor (KNN) Classifier and the Support Vector Machine (SVM) Classifier ($p<0.05$ Independent sample test). X-axis: Support Vector Machine (SVM) Classifier accuracy rate vs K-Nearest Neighbor (KNN) Classifier Y-axis: Mean of accuracy rate, for identification of keywords ± 1 SD with 95 % CI. Table.1 shows the Evaluation Metrics of Comparison of K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) Classifier. The accuracy rate of K-Nearest Neighbor (KNN) is 87.73 and Support Vector Machine (SVM) is 93.047. In all aspects of parameters Support Vector Machine (SVM) provides better performance compared with the K-Nearest Neighbor (KNN) of predicting Spam comments on YouTube videos with improved accuracy rate. Table.2 shows the statistical calculation such as Mean, standard deviation and standard error Mean for K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). The accuracy rate parameter used in the t-test. The mean accuracy rate of K-Nearest Neighbor (KNN) is 87.73 and Support Vector Machine (SVM) is 93.047. The Standard Deviation of K-Nearest Neighbor (KNN) is 1.82738 and Support Vector Machine (SVM) is 0.23323. The Standard Error Mean of K-Nearest Neighbor (KNN) is 0.93827 and Support Vector Machine (SVM) is 0.18343. Table.3 displays the statistical calculations for independent samples tested between K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). The sig. the signal to noise ratio is 0.022. Independent samples T-test is applied for comparison of K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) with the confidence interval as 95% and level of significance as 0.12323. This independent sample test consists of significance as 0.001, significance (2-tailed), Mean difference, standard error difference, and lower and upper interval difference.

4. Discussion

A test is conducted between K-Nearest Neighbor (KNN) and Innovative Support Vector Machine

(SVM). Classification is a process to classify the YouTube comment into Ham or Spam message. The result shows that SVM produces higher accuracy than KNN. As features identified from the literature review, various features may be extracted from YouTube classification purposes. Besides, the data already consists of two classes where the classes are "spam" and "ham". Thus, easy to choose features that certainly label as spam. YouTube comments may contain hyperlinks, text, uppercase and lowercase characters. However, those uppercase characters do not exist after the preprocessing phase. After preprocessing, this study decides to use keywords as feature selection. The Confusion Matrix is a deep learning visual exploration method. The results of the prediction class are represented in the Confusion Matrix columns, while the results of the actual class are represented in rows. This matrix incorporates all raw data in respect of the classification model considerations in the specified data collection. Determining how accurate a model is. It is a square matrix with lines representing the actual category of cases and columns representing their expected category. The confusion matrix is a 2 x 2 matrix that reports the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) when working with a Binary classification mission. Precision, recall, and F-measure, which is widely used in text mining and machine learning communities, has been used to test algorithms. True positive (TP - items with the appropriate label as class), false positive (FP - false labeled items), false (FN - items labeled incorrectly classified), and true negative (TN - improperly labeled items) are four types of classified items (TN - well-classified items). Recall is determined using the following formula given the number of true benefits and false positives

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (1)$$

Many researchers had been studying the classification of YouTube comments as spam or ham by using machine learning. Kantchelian et al.(Kantchelian et al. 2012) developed a logistic regression based entropy rate. Aziz et al.(Aziz et al. 2018) discussed the performance comparison of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Alias et al.(Alias, Foozy, and Ramli 2019) used six classifiers 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 machines 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. Aiyar dan Shetty(Aiyar and Shetty 2018) 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 combating spam comments. Al-Zoubi et al.(Al-Zoubi et al. 2018) 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 and AUC. In the proposed work, we include the most important Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) models for predicting the spam comments. The accuracy rate of the Support Vector Machine (SVM) is 93.047% compared with K-Nearest Neighbor (KNN) that has an accuracy rate of 87.73%. The structure of the modified information box takes into account the intermediate, minor and major attributes, interquartile access and malfunction (focus beyond the upper and lower extremities). In addition, the application includes the division of the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) Model to determine whether it is possible for predicting the spam comments. The findings indicated most of the analyzed methods of classification are for filtering spam comments on YouTube. Most of them were able to provide small or no blocked ham rates and it was also possible to achieve accuracy rates that were more than 90 per cent. The limitation of the proposed work is that the features do not cover the user perception of a video, since our dataset did not cover this information. The regression analyses on the relationship of comment types and user ratings would certainly reveal further interesting results if we included the video content perception as another independent variable. Further

research will be necessary to analyze these correlations.

5. Conclusion

In this research, the development of a Youtube spam comment detection framework by using machine learning techniques has been done. It is important to improve security since the Internet nowadays indicates the security issues. The proposed model exhibits the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM), in which the Support Vector Machine (SVM) has the highest values. The accuracy Rate of Support Vector Machine (SVM) is 93.047 is higher compared with K-Nearest Neighbor (KNN) that has an accuracy rate of 87.73 in analysis of detecting Spam comments on YouTube videos with improved accuracy rate.

Declaration

Conflicts of Interest

No conflict of interest in this manuscript

Authors Contributions

Author KR was involved in data collection, data analysis & manuscript writing. Author MVP was involved in conceptualization, data validation, and critical review of manuscripts.

Acknowledgment

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for successfully carrying out this work.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

1. Chipontime Technologies Pvt. Ltd. Bangalore.
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

6. References

Aiyar, Shreyas, and Nisha P. Shetty. 2018. "N-Gram Assisted Youtube Spam Comment Detection." Procedia Computer Science 132 (January): 174–82.

Alberto, Tulio C., Johannes V. Lochter, and Tiago A. Almeida. 2015. "TubeSpam: Comment Spam Filtering on YouTube." 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). https://doi.org/10.1109/icmla.2015.37.

Alberto, Túlio C., Johannes V. Lochter, and Tiago A. Almeida. 2015. "TubeSpam: Comment Spam Filtering on YouTube." In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), 138–43.

Alias, N., C. F. M. Foozy, and S. N. Ramli. 2019. "Video Spam Comment Features Selection Using Machine Learning Techniques." Indones. J. Electr. https://pdfs.semanticscholar.org/482c/9c66a70d72bed284d59740d6a26322d51b6c.pdf.

Al-Zoubi, Ala' M., Hossam Faris, Ja'far Alqatawna, and Mohammad A. Hassonah. 2018. "Evolving Support Vector Machines Using Whale Optimization Algorithm for Spam Profiles Detection on Online Social Networks in Different Lingual Contexts." Knowledge-Based Systems. https://doi.org/10.1016/j.knosys.2018.04.025.

Aziz, Aqliima, Cik Feresa Mohd Foozy, Palaniappan Shamala, and Zurinah Suradi. 2018. "YouTube Spam Comment Detection Using Support Vector Machine and K–Nearest Neighbor." Indonesian Journal of Electrical Engineering and Computer Science. https://doi.org/10.11591/ijeecs.v12.i2.pp612-619.

Chowdury, Rashid, Md Nuruddin Monsur Adnan, G. A. N. Mahmud, and Rashedur M. Rahman. 2013. "A Data Mining Based Spam Detection System for YouTube." In Eighth International Conference on Digital Information Management (ICDIM 2013), 373–78.

Figueiredo, Flavio, Fabrício Benevenuto, and Jussara M. Almeida. 2011. "The Tube over Time: Characterizing Popularity Growth of Youtube Videos." In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 745–54. WSDM '11. New York, NY, USA: Association for Computing Machinery.

Gaye, Babacar, Dezheng Zhang, and Aziguli Wulamu. 2021. "Improvement of Support Vector Machine Algorithm in Big Data Background." Mathematical Problems in Engineering 2021 (June). https://doi.org/10.1155/2021/5594899.

Goyal, Saumya, R. K. Chauhan, and Shabnam

Parveen. 2016. "Spam Detection Using KNN and Decision Tree Mechanism in Social Network." In 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), 522–26.

Grabowicz, Przemyslaw A., José J. Ramasco, Esteban Moro, Josep M. Pujol, and Victor M. Eguiluz. 2012. "Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media." PloS One 7 (1): e29358.

Hilbe, Joseph M. 2004. "A Review of SPSS 12.01, Part 2." The American Statistician 58 (2): 168–71.

Ingale, Mayur, and K. S. Bhagat2& P. M. Mahajan. n.d. "Image Segmentation Using Gabor Filter And Wavelet Transform." International Journal Of Electrical And Electronics Engineering Research (IJEEER) ISSN, 25–28.

Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." Energy. https://doi.org/10.1016/j.energy.2022.123709.

Jayakodi, Santhoshkumar, Rajeshkumar Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." Environmental Research 212 (Pt A): 113153.

Kanodia, Simran, Rachna Sasheendran, and Vinod Pathari. 2018. "A Novel Approach for Youtube Video Spam Detection Using Markov Decision Process." In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 60–66. ieeexplore.ieee.org.

Kantchelian, Alex, Justin Ma, Ling Huang, Sadia Afroz, Anthony Joseph, and J. D. Tygar. 2012. "Robust Detection of Comment Spam Using Entropy Rate." In Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence, 59–70. AISec '12. New York, NY, USA: Association for Computing Machinery.

Kiran, P. S. 2015. "Detecting Spammers in YouTube: A Study to Find Spam Content in a Video Platform." IOSR Journal of Engineering (IOSRJEN) 5 (7): 26–30.

Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." Computers and Electrical Engineering. https://doi.org/10.1016/j.compeleceng.2022.107902.

Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using NCI Database." Korean Journal of Chemical Engineering. https://doi.org/10.1007/s11814-022-1120-5.

Lichman, Moshe, and Others. 2013. "UCI Machine Learning Repository, 2013." URL Http://archive. Ics. Uci. Edu/ml 40.

Manwar, Supriya Ramhari, P. Lambhate, and J. Patil. 2017. "Classification Methods for Spam Detection In Online Social Network." https://www.academia.edu/download/54172935/IRJET-V4I7230.pdf.

Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh, Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022. "Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." Environmental Research. https://doi.org/10.1016/j.envres.2022.113136.

Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." Carbohydrate Polymers 287 (July): 119349.

Rădulescu, Cristina, Mihaela Dinsoreanu, and Rodica Potolea. 2014. "Identification of Spam Comments Using Natural Language Processing Techniques." In 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP), 29–35.

Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." Chemosphere 298

(July): 134341.

Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." Chemosphere 298 (July): 134265.

Sharmin, Sadia, and Zakia Zaman. 2017. "Spam Detection in Social Media Employing Machine Learning Tool for Text Mining." In 2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), 137–42. ieeexplore.ieee.org.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R. Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-022-06636-5.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." Environmental Research. https://doi.org/10.1016/j.envres.2022.113114.

Yusof, Yuhanis, and Omar Hadeb Sadoon. 2017. "Detecting Video Spammers in YouTube Social Media." In , 228–34. repo.uum.edu.my.

Zhang, Qiang, Chenwei Liu, Shangru Zhong, and Kai Lei. 2017. "Spam Comments Detection with Self-Extensible Dictionary and Text-Based Features." In 2017 IEEE Symposium on Computers and Communications (ISCC), 1225–30. ieeexplore.ieee.org.

Tables and Figures

Table 1: Comparison of K-Nearest Neighbor (KNN) and Innovative Support Vector Machine (SVM) Classifier for predicting the spam comments in YouTube videos with improved accuracy rate. The accuracy rate of K-Nearest Neighbor (KNN) is 87.73 and Innovative Support Vector Machine (SVM) has 93.047

SI.No.z	Test Size	Accuracy Rate	
		K-Nearest Neighbor	Support vector Machine
1	Test 1	84.32	86.32
2	Test2	84.13	87.24
3	Test3	84.71	88.64
4	Test4	85.10	89.67
5	Test5	85.35	89.84
6	Test6	85.23	90.76
7	Test7	86.01	91.14
8	Test8	86.01	91.34
9	Test9	86.12	92.12
10	Test10	86.23	92.43

Table .2. The statistical calculation such as Mean, standard deviation and standard error Mean for K-Nearest Neighbor (KNN) and Innovative Support Vector Machine (SVM). The accuracy rate parameter used in the t-test. The mean accuracy rate of K-Nearest Neighbor (KNN) is 87.73 and Support Vector Machine (SVM) is 93.047. The Standard Deviation of K-Nearest Neighbor (KNN) is 1.82738 and Support Vector Machine (SVM) is 0.23323. The Standard Error Mean of K-Nearest Neighbor (KNN) is 0.93827 and Support Vector Machine (SVM) is 0.18343.

Group		**N**	**Mean**	**Standard Deviation**	**Standard Error Mean**
Accuracy	Decision Tree	10	83.047	0.23323	0.18343
	K-Nearest Neighbor (Knn)	10	87.73	1.82738	0.93827

Table. 3: The statistical calculations for independent samples test between K-Nearest Neighbor (KNN) and Innovative Support Vector Machine (SVM). The sig. for signal to noise ratio is 0.022. Independent samples T-test is applied for comparison of K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) with the confidence interval as 95% and level of significance as 0.12323. This independent sample test consists of significance as 0.001, significance (2-tailed), Mean difference, standard error difference, and lower and upper interval difference.

Group		**Levene,s Test for Equality Of Variances**		**T-test for Equality of Mean**						
		F	**sig.**	**t**	**df**	**sig(2-tailed)**	**Median difference**	**Std Error Difference**	**95% Confidence Interval (Lower)**	**95% Confidence Interva1 (Upper)l**
Accurac	Equal Variances	7.033	0.022	2.700	18	.010	17.978	0.8228	16.4918	12.9748

y	assumed									
	Equal Variances assumed			2.700	17.978	.010	17.978	0.8228	16.4918	11.9748

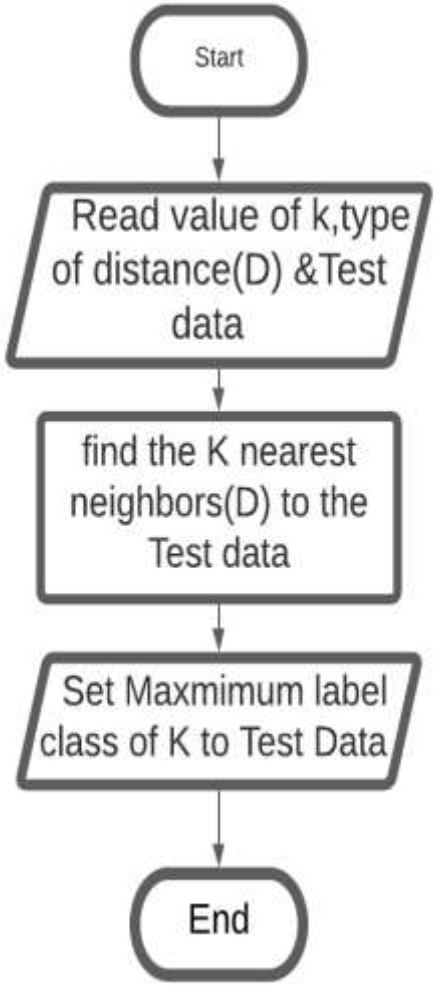

Fig .1. Flow Chart of K-Nearest Neighbor

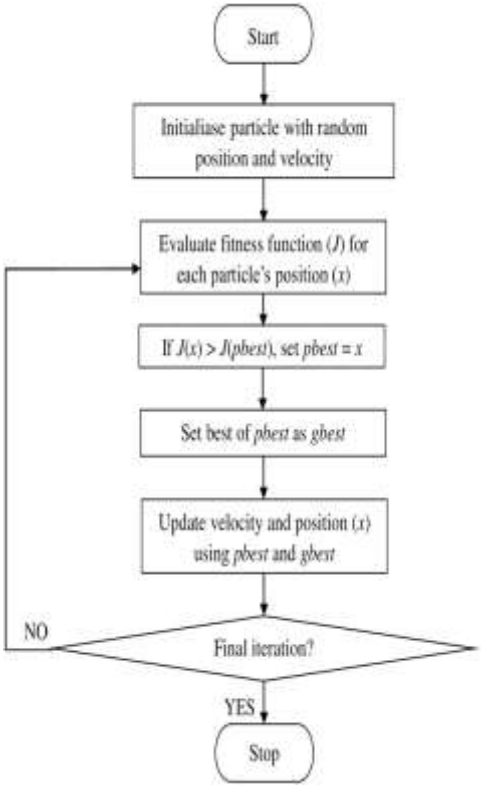

Fig. 2. Flow Chart of Support Vector Machine

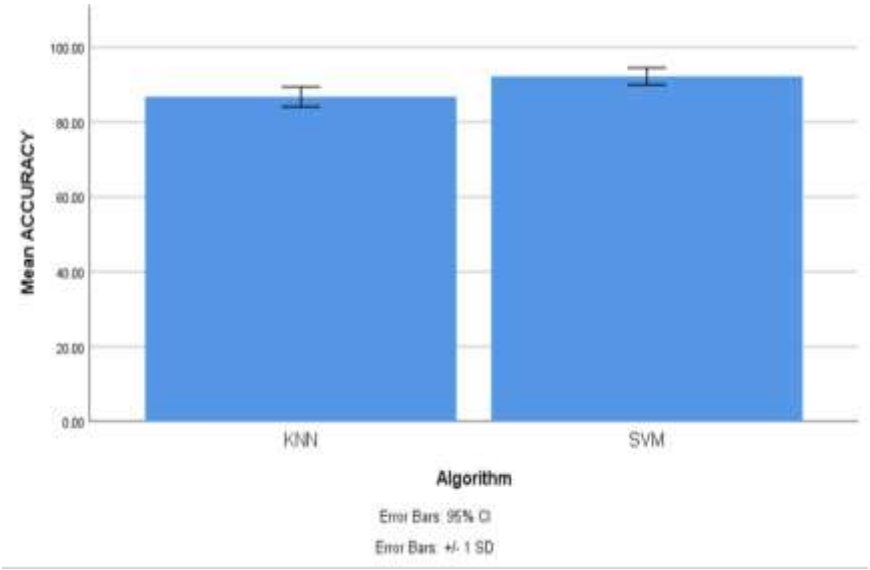

Fig. 3. Bar graph between KNN and Innovative Support Vector Machine (SVM). Comparison of KNN algorithm and SVM in terms of mean accuracy. The mean accuracy of KNN is better than NB and the standard deviation of KNN is slightly better than SVM. X-Axis: KNN vs SVM Y-Axis: Mean accuracy of detection ± 1 SD with CI of 95%.