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.2023.52930

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*

# Social Media Spam Comments Detection Analysis using Machine Learning

Vaibhav Ambhore[1], Adwait Garud[2], Harshal Lande[3], Pratap Rede[4]
*D. Y. Patil College of Engineering Ambi, Pune-410507*

***Abstract:*** ***The results published by Google's famous video distribution platform Social Media are attracting more and more users every day. However, this success has also attracted many malicious users who try to promote their videos or infect them with viruses and malware. As we know Social Media has a tool to measure comments so spam is increasing very fast which is why the owner's comment section is closed. Creating a classification system for automatic spam filtering is very difficult because words are short and often contain slang words, symbols, and abbreviations. In this article, we consider several best practices for detecting and analyzing spam messages.***

***Analysis of the results showed that decision trees, logistic regression, Bernoulli Naive Bayes, Random Forest, Linear and Gaussian SVMs were equal to the highest values, with 99.9% confidence. That's why it's important to find a way to check video comments and ads before they're viewed by unsuspecting users.***

***Keywords: Machine learning, linear regression, random forests, prediction.***

## I. INTRODUCTION

In an era dominated by social media, platforms such as Facebook, Instagram, Twitter and YouTube have become hotspots where users communicate, discuss and express themselves. However, the growing popularity of these platforms has also attracted many spam comments, making it difficult to maintain a positive and engaging online presence. In order to solve this problem, the use of machine learning techniques in the detection of spam messages has attracted attention.

Machine learning algorithms have proven useful in many areas, including natural language processing (NLP) and pattern recognition. Using the power of machine learning, researchers and developers have sought to create efficient and accurate models that can identify and filter spam on social media platforms.

The purpose of this analysis is to explore the use of machine learning techniques in detecting spam comments. By analysing the characteristics and patterns of spam messages, we aim to build a model that can effectively analyse and filter out spam content.

Analyzing spam comments involves processing a lot of user-generated data, which presents unique challenges. Legislation based on methods or concepts is briefly compared in relation to the evolution of spam technology. Machine learning, on the other hand, is more flexible and scalable.

In this review, we will explore various machine learning algorithms such as support vector machine (SVM), decision trees, random forests, and deep learning models to generate the proposed spam detection. We will leverage the power of NLP techniques, including content extraction, text classification and sentiment analysis, to train and evaluate our models.

Data used for training and assessment may contain comments, including spam and suggestions from social media platforms. Using a suitable pre-processing method, we will clean the raw data and transform the model into a suitable one for training.

Evaluation of the design will be based on criteria such as accuracy, precision, recall and F1 score to evaluate their performance in distinguishing spam from legitimate messages.

As spam analysis seems to be in the minority, we will also explore strategies for dealing with data inconsistencies.

The results of this analysis will give you a better idea of the performance of machine learning algorithms in detecting spam on social media platforms. By creating the right model, we can help platform administrators and moderators get spam filters, thereby improving user experience and reducing the occurrence of inappropriate content language.

Overall, this review aims to contribute to research and techniques that use machine learning to combat spam messages on social media platforms. By leveraging the power of Naive Bayes algorithms and NLP technology, we strive to create a safe and effective online experience for our users.

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue V May 2023- Available at www.ijraset.com*

## II. EXISTING SYSTEM

Over the years, machine learning has been at the forefront of a lot of research on online spam or spammer or spam account detection.

Researchers used different techniques when designing spam detection methods using machine learning techniques. While researchers work to detect spam, spammers work hard to avoid detection. Spam comments are defined as comments that contain advertisements or appear to be unrelated to the content of a video. Many algorithms have been used before to identify spam, for example CNN (Convolutional Neural Network).

We use Naive Bayes classifier and Support Vector Machine classifier (SVM) in the planning process. Naive Bayes is used to calculate the number of possibilities, which means that using spam money as an example results in spam money. SVMs is used for data classification.

## III. PROBLEM STATEMENT

To develop a system which is capable of performing following actions:

1) *Training System on a Test Dataset in order to Recognize spam Comments:* The training data is the largest (in terms of size) subset of the original data used to train or fit the machine learning model. First, the training data is fed into machine learning algorithms, allowing them to learn how to predict a task.
2) *Test System for Test Data:* After training the model with the training data, it's time to evaluate the model with the test data. This information evaluates the performance of the model and enables the model to adapt to new or unknown data. Test data is another primary dataset that is independent of the training data.
3) *Test Based on Real Spam Files:* The active spam testing system receives spam specifications from testers, which will be pulled into the search engine for testing. This test can generate information to help track the signs of spam to be reported.

## IV. IMPLEMENTATION STRATEGY

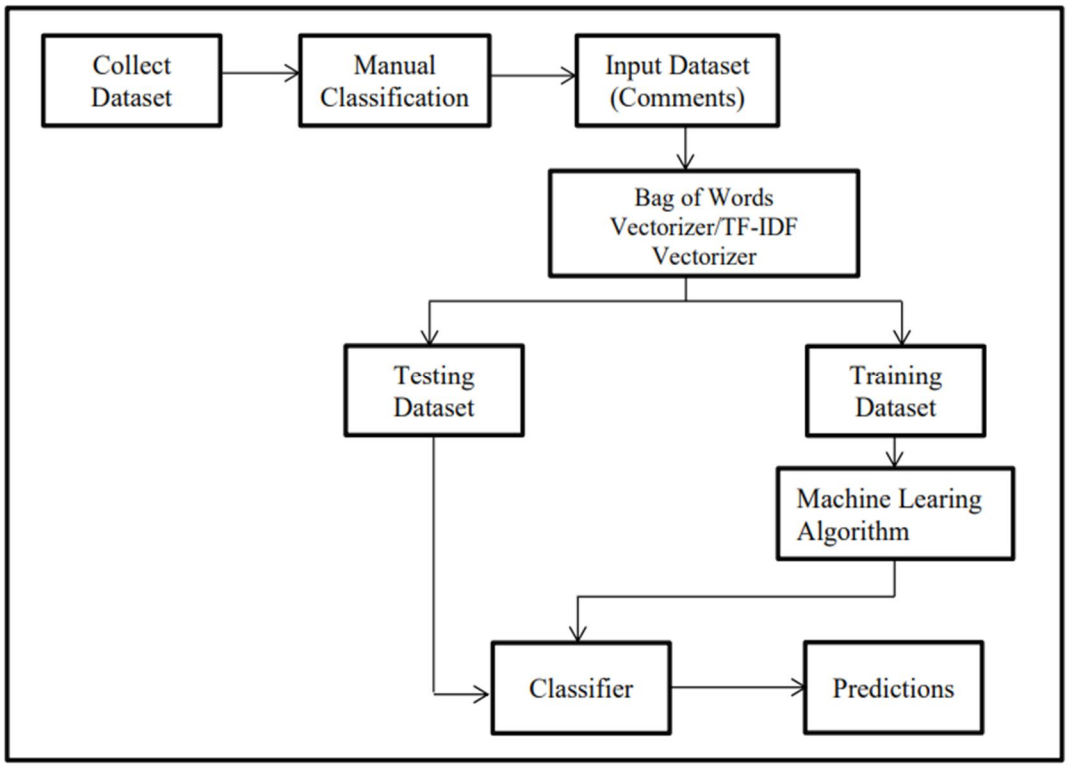
1) Create an infographic using 4 out of 5 Social Media files.
2) Since the dataset is a data file, attributes must be provided by the (CONTENT) field.
3) It is important to bring properties from the document only to the "CONTENT" field in the document.
4) Remove unwanted special characters (ï»¿) from the response.
5) Build a spam dictionary of words commonly found in spam by looking at spam patterns.
6) Use the spam dictionary to count spam messages in messages.
7) To check if the comment contains strings "http “,"www" or ".com" string which represent promotions and could be SPAM and set IS_HTTP=1 else
8) To calculate the ratio of spam to the number of words in the message, first remove the English words "STOPWORDS" such as (I, me, etc.) from the replies
9) Find out the length of the message, because long messages are usually spam9. Count the number of words in the review after removing the stop words.
10) Calculate the ratio of spam messages to all messages.
11) Use Naive Bayes classifier on test data and validate with test data

## V. PROPOSED SYSTEM

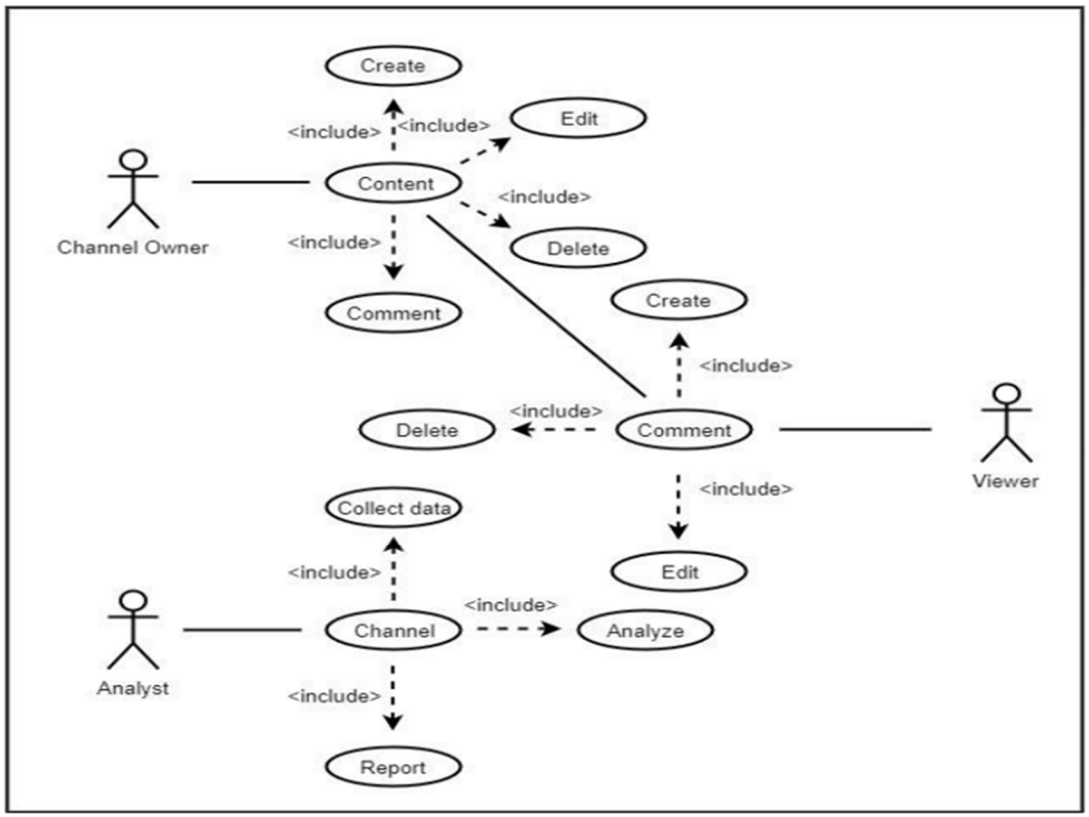
We try to detect spam messages using some traditional heuristics, such as the traditional machine learning algorithm Naive Bayes, as well as N-Grams, which has proven effective in searching and after fighting spam. We have collected and created five databases containing real, public and uncensored data directly from Social Media via API7. We've selected five of the ten most-watched YouTube videos. Each sample represents the word spoken in the speech of each selected video. No previous procedures were applied.

Each sample is then tagged as spam or legal (raw) using a common tool called tagging created for this purpose. The model is associated with metadata such as the stored name and publication date.




## VI. USE CASE VIEW



## VII. ALGORITHMS USED

### A. Naive Bayes

Naive Bayes is a classification algorithm among machine learning algorithms. It is used only to solve the problem of classification of texts with high educational content. It is a probability-based classifier. Probability includes the probability that an event will occur based on experience. The algorithm is known for its simplicity and efficiency.

Bayes' theorem, also known as Bayes' rule or Bayes' rule, is used to determine the probability of a previously informed hypothesis. Depends on what it is.

The formula for Bayes' theorem is given as: $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P(A) is Prior Probability: Probability of hypothesis before observing the evidence.

P(B) is Marginal Probability: Probability of Evidence.

### B. C4.5

C4.5 is a pedigree machine learning algorithm. The algorithm uses data entropy to build a tree from the given data. We deliberately analyze and classify data from large collections and yield similar results.

C4.5 is an extension of the ID3 algorithm. The algorithm uses the obtained data to divide the sample data into a series of subsets, choosing the data attributes for each node, and the highest value is determined. This process reverts to the smallest sublist on the other hand, if you are dealing with environmental data such as weather data or historical crop data, you can use RNN-based models such as Long Short Term Memory (LSTM) or Gated Recurrent Unit. (GRU)) may be more appropriate. RNNs are designed to capture trends in data, making them ideal for time series analysis tasks.

## VIII. CONCLUSION

Social media has become popular, creating an opportunity for spammers to spread unwanted messages.

Previously, some machine learning algorithms were used for this detection. In the planning process, we also use advanced machine learning algorithms with advanced features and use them to compare the results of various algorithms. We create features based on features derived from user profiles and the content they share. Based on the test results, it can be assumed that existing products widely used in the data mining community can use these features to identify spammers.

## IX. PROJECT SCOPE AND FEASIBILITY

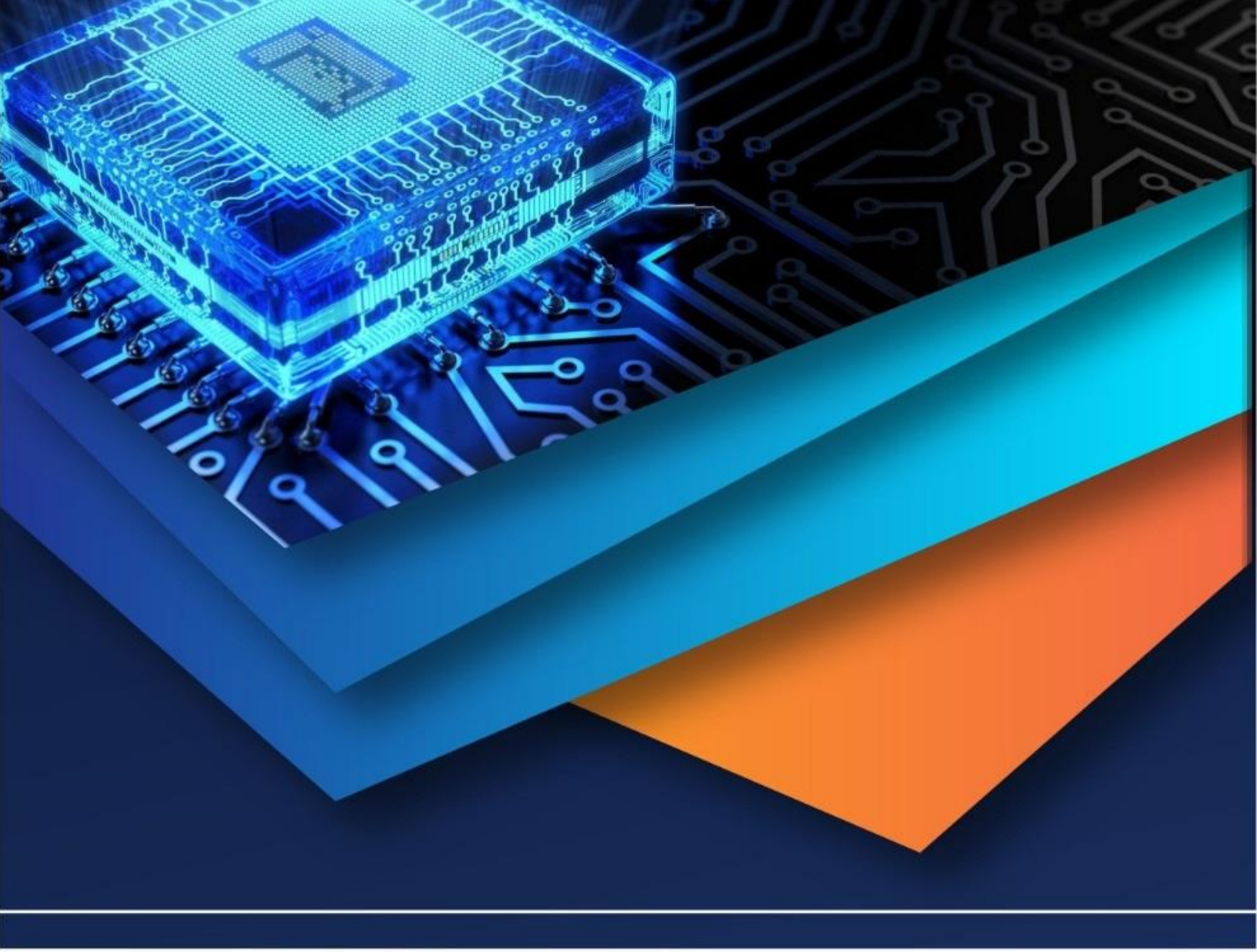
*1)* Social Media is a large video sharing network with a large user base. It is not possible to follow such a large-scale project at our level. However, the project was carried out on some limits. This work can be extended to more users and larger data.

*2)* A feasibility study is an evaluation of the effectiveness of a proposal or process.

The project needs information that can be completed on time. The information needed by the Theis project has been posted in the public domain. The project is possible because all the components required for the project are available

## REFERENCES

[1] K. S. Adewole, N. B. Anuar, A. Kamsin, K. D. Varathan and S. A. Razak, "Malicious accounts: Dark of the social networks", Elsevier, 2017, pp. 41-67.
[2] A. Gupta and R. Kaushal, "Improving Spam Detection in Online Social Networks", IEEE, 2015.
[3] M. Verma, Divya, S. Sofat, "Techniques to Detect Spammers in Twitter – A Survey", International Journal of Computer Applications, January 2014, Vol. 85, No. 10, pp. 27-32.
[4] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou and G. Min, "Statistical Features- Based Real- Time Detection of Drifted Twitter Spam", IEEE Transactions, April 2017, pp.914-925



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY