**MPSTME NMIMS, Shirpur**

**Final Project Report**

**on**

**Youtube Spam Comment Detection**

Team members:

Pranav Kolhe

Guided by:

Prof. Dhananjay Joshi

Industry Mentor:

**Table of Contents**

**Executive Summary**

1. **Background……….. 1-2**
   1. **Scope…………….1**
   2. **Overview……….. 2**
2. **System**
   1. **Requirements**
      1. Functional requirements……………3
      2. User requirements…………………3
      3. Non-Functional requirements………3
   2. **Data Design…………4**
   3. **Implementation…..5**
3. **Snapshots of the Project………6-11**
4. **Literature Survey…….12**
5. **References……………12**

**Executive Summary**

YouTube, the world’s largest video sharing site, was founded in 2005 and acquired by Google in 2006. YouTube has grown tremendously as a video content platform, with the recent shift in online content to video.

At present, more than 400 hours of video are uploaded and 4.5 million videos are watched every minute on YouTube. It is easy for users to watch and upload videos without any restrictions.

This great accessibility has increased the number of personal media, and some of them have become online influencers.

YouTube creators can monetize if they have more than 1,000 subscribers and 4,000 hours of watch time for the last 12 months. Accordingly, spam comments are being created to promote their channels or videos in popular videos.

Some creators closed the comment function due to aggression such as political comments, abusive speech, or derogatory comments not related to their videos.

YouTube has its own spam filtering system, though there are still spam comments that are not being caught.

With the growth of Youtube, there is a growth of Spam Comments on it too.

Youtubers don’t have the means to address that, but it can be addressed through data science and coding.

Through this project, I will try to do the same and hope to improve on the previously implemented projects.

**Scope of the document:**

The scope of the project is to deal with the Spam Comment that are present on Youtube Videos.

A lot of Youtube channels deal with spam comments. Them not having the technical abilities to deal with it, they can’t help but leave those comments up.

Through this project, I plan to help Youtube and Youtube users classify whether the comment is a scam or not using the code, and if it is, they can simply delete the comment or block the person who is commenting it on their videos.

**Intended Audience:**

This project would be of help for people from different fields:

* Youtubers: Youtubers who regularly upload videos and face this issue can use this system to avoid and tackle this problem.
* Employees: All the complaints published to the employees regarding the spam comments can be solved through this project.

**System Overview:**

Depending on how it's set up, the way a youtube spam comment detection system works can be different, but here are some general steps:

* Problem Definition: Define the problem and the requirements for the youtube Spam Comment detection system. Identify the key features and performance metrics, such as accuracy, precision, and recall.
* Data Collection: Collect a dataset of comments on youtube videos with and without spam comments. The dataset should be diverse and representative of the target population.
* Data Preprocessing: Preprocess the dataset by choosing the relevant data to focus on the spam comments’ features.
* Model Selection: Select a suitable machine learning algorithm for the youtube Spam Comment detection task, such as Support Vector Machines (SVMs), Random Forests, or K-Nearest Neighbor (k-NN). Train and validate the model on the preprocessed dataset, using cross-validation techniques to evaluate its performance.
* Model Optimization: Optimize the model by tuning its hyperparameters, such as learning rate, regularization strength, and batch size. Use grid search or other optimization techniques to find the best combination of hyperparameters.
* Testing and Evaluation: Test and evaluate the youtube Spam Comment detection system on a new dataset of comments. Use performance metrics such as accuracy, precision, and recall to evaluate its performance. Iterate on the design and implementation of the system based on the results of testing and evaluation.

**Problem Statement:**

The popularity of YouTube not only attracted genuine viewers but spammers as well. As a result, there is an increase in unwanted spam videos and comments. Here comes the importance of an AI-based YouTube spam comment detection model.

In this AI project, you will be focusing on text and words and classify internet comments as spam or not spam. The spam detection model can be accomplished by using bagof-words and random forest techniques.

You can also predict positive and negative reviews with the Word2Vec approach and the k-nearest neighbor classifier in addition to spam detection.

Functional Requirements

4.1 Comment Analysis:

The system should analyze comments on YouTube videos, looking for characteristics that indicate spam, such as irrelevant content or excessive links.

The system should use machine learning algorithms to classify comments as spam or not spam.

The system should store the results of the analysis in a database.

4.2 Comment Filtering:

The system should filter out spam comments, preventing them from being displayed on YouTube videos.

The system should allow administrators to adjust the level of filtering based on their preferences.

4.3 Reporting:

The system should provide a reporting system that allows users to flag spam comments for review.

The system should allow administrators to review flagged comments and take appropriate action.

User Requirements:

User-friendly interface: The system should have a user-friendly interface that is easy to navigate and understand.

Reporting system: The system should have a reporting system that allows users to flag spam comments for review.

Transparency: The system should be transparent about how it works and what rules it uses to detect spam comments.

Customizability: The system should allow users to customize the level of filtering they want to apply to comments, including the ability to whitelist or blacklist certain words or phrases.

Privacy: The system should protect user privacy by not storing or sharing any personal information.

System Requirements:

Machine Learning algorithms: The system should use machine learning algorithms to detect spam comments.

Real-time detection: The system should be able to detect spam comments in real-time.

Scalability: The system should be scalable to handle large volumes of comments.

High accuracy: The system should have a high accuracy rate in detecting spam comments.

Customizability: The system should allow administrators to customize filtering rules for spam comments.

Security: The system should be secure to prevent unauthorized access and protect user data.

Compatibility: The system should be compatible with different platforms and devices.

Reliability: The system should be reliable and available 24/7.

Performance: The system should have high performance to ensure real-time detection of spam comments.

Non-functional Requirements

5.1 Usability: The system should be user-friendly and easy to navigate.

5.2 Performance: The system should have high performance and be able to handle large volumes of comments in real-time.

5.3 Reliability: The system should be reliable and available 24/7.

5.4 Security: The system should be secure and protect user data from unauthorized access.

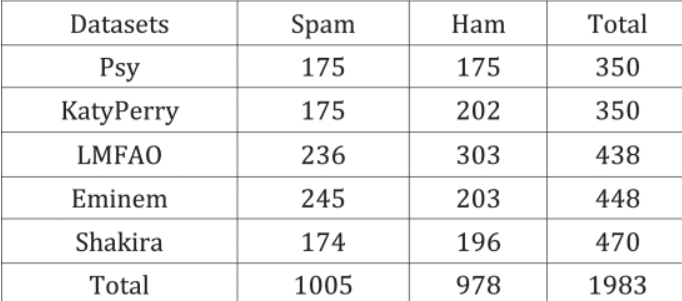
**Data Design:**

It is a public set of comments collected for spam research. It has five datasets composed by 1,956 real messages extracted from five videos that were among the 10 most viewed on the collection period.

The collection is composed by one CSV file per dataset, where each line has the following attributes:

* COMMENT\_ID,AUTHOR,DATE,CONTENT,TAG

They contain YouTube ID, comment author, date, comment content, and labeled class (0: Ham or 1: Spam). We only use comment content and labeled class.

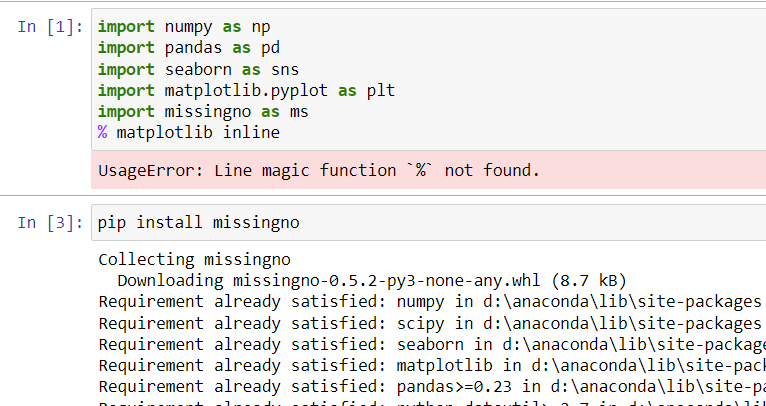


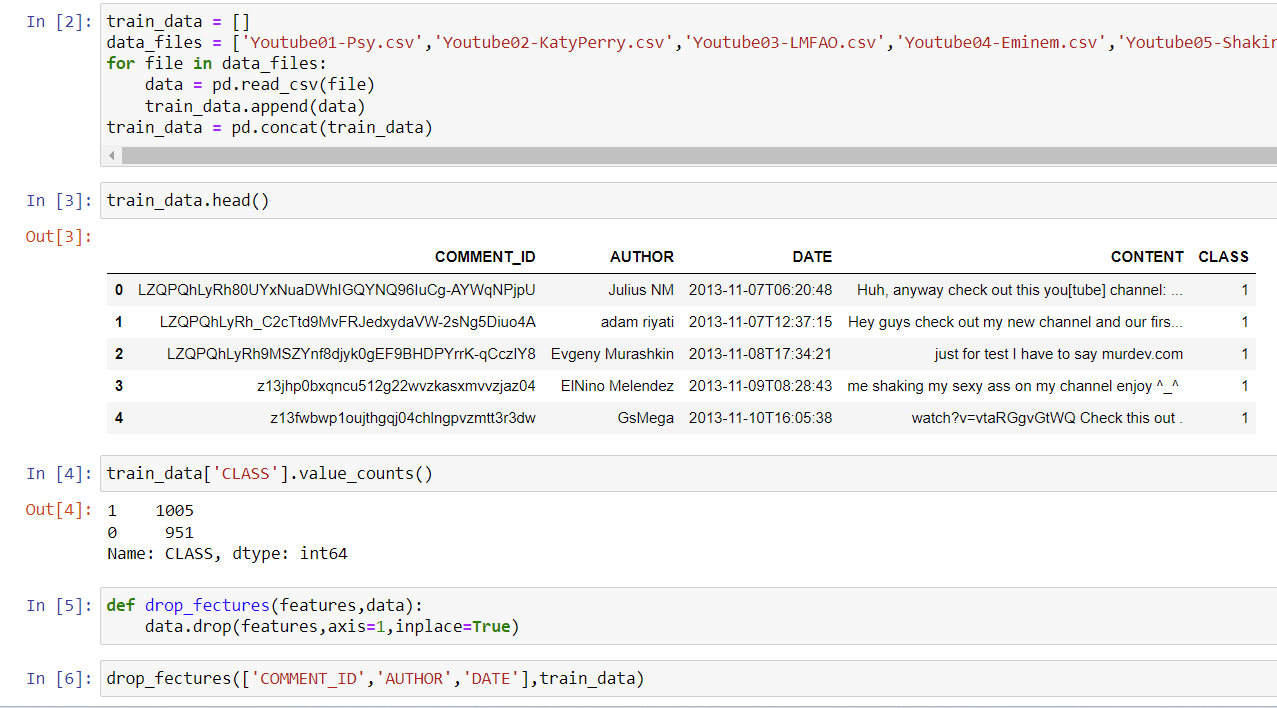
**Implementation:**

Depending on how it's set up, the way a youtube spam comment detection system works can be different, but here are some general steps:

* Problem Definition: Define the problem and the requirements for the youtube Spam Comment detection system. Identify the key features and performance metrics, such as accuracy, precision, and recall.
* Data Collection: Collect a dataset of comments on youtube videos with and without spam comments. The dataset should be diverse and representative of the target population.
* Data Preprocessing: Preprocess the dataset by choosing the relevant data to focus on the spam comments’ features.
* Model Selection: Select a suitable machine learning algorithm for the youtube Spam Comment detection task, such as Support Vector Machines (SVMs), Random Forests, or K-Nearest Neighbor (k-NN). Train and validate the model on the preprocessed dataset, using cross-validation techniques to evaluate its performance.
* Model Optimization: Optimize the model by tuning its hyperparameters, such as learning rate, regularization strength, and batch size. Use grid search or other optimization techniques to find the best combination of hyperparameters.
* Testing and Evaluation: Test and evaluate the youtube Spam Comment detection system on a new dataset of comments. Use performance metrics such as accuracy, precision, and recall to evaluate its performance. Iterate on the design and implementation of the system based on the results of testing and evaluation.

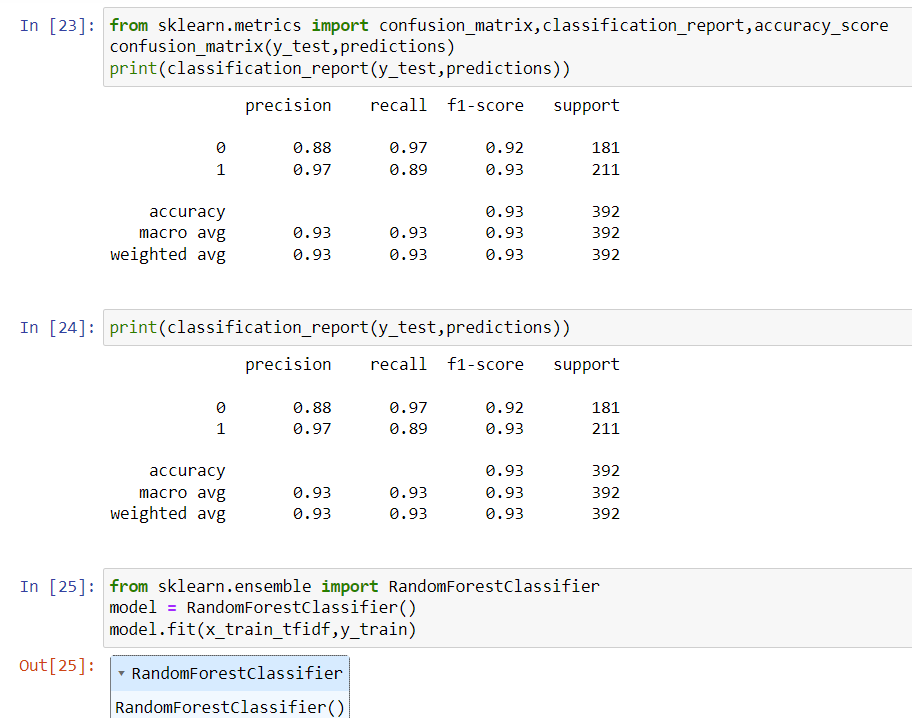
**Snapshots of Code:**

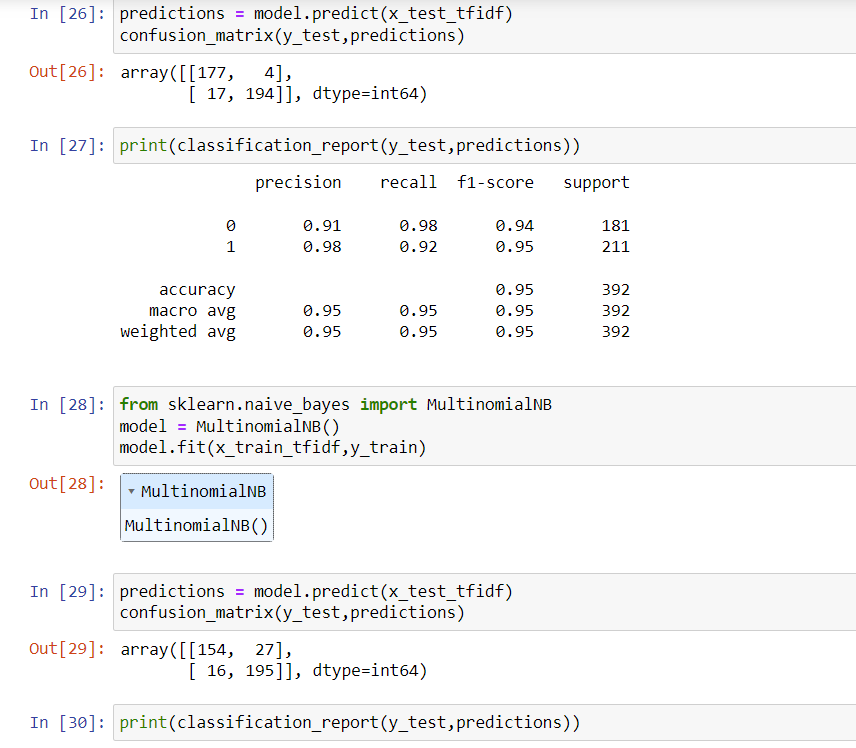


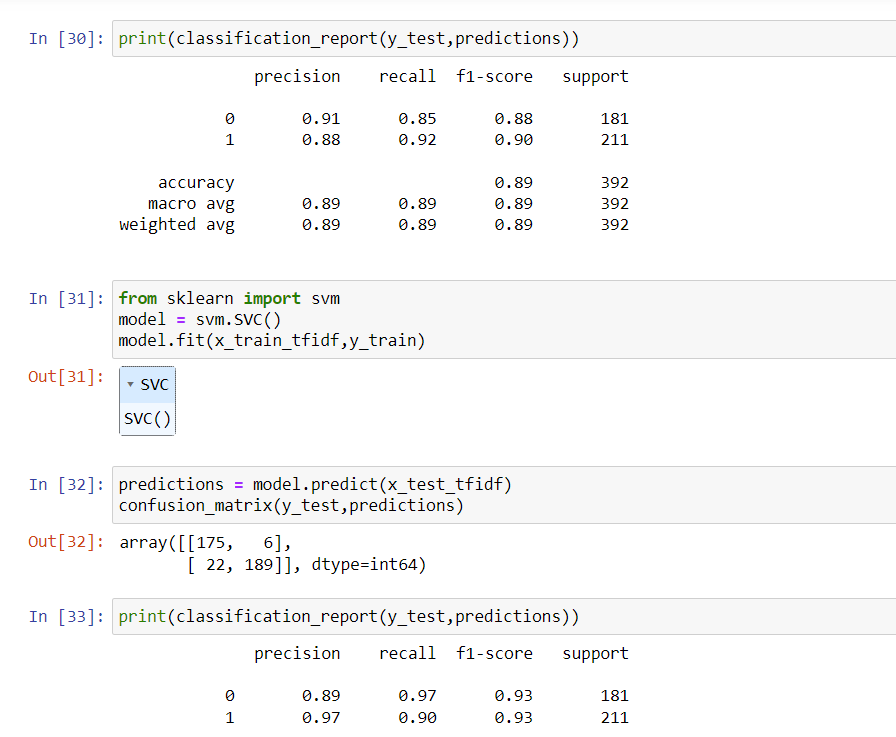


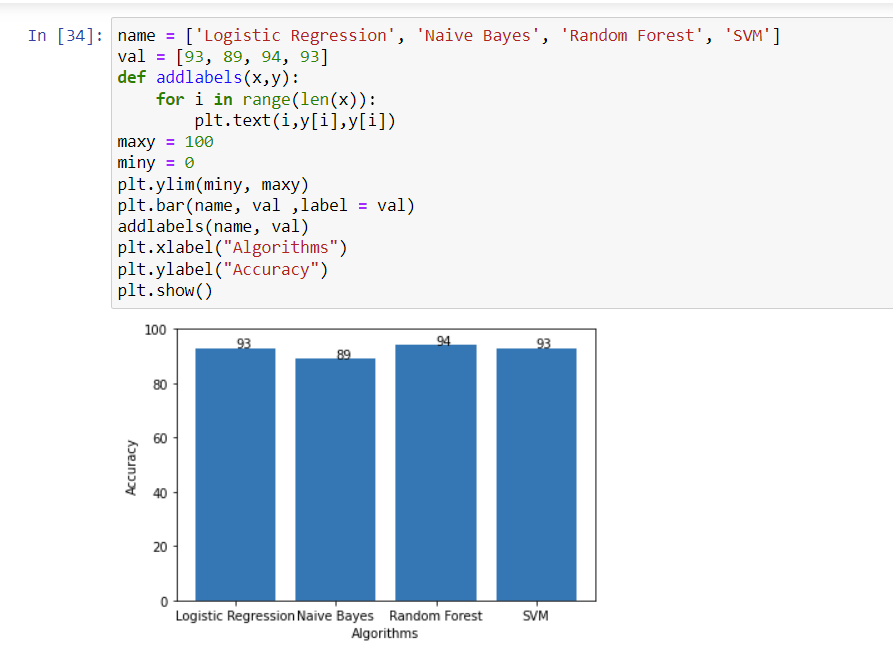












# Literature Survey

Here are three papers related to youtube Spam Comment detection, along with a brief summary of each and their contributions:

* "Youtube Spam Comments Detection" by Oviya Selvaraj, Anuradha Konatham, Dr. Paavai Anand. (2020)

In this paper, the spam comments detection methodology has been carried out using four Artificial Intelligence estimations – Logistic Regression, Ada Boost, Decision Tree and Random Forest. With the use of Neural Network, we can achieve an exactness of 91.65% and beat the present course of action by around 18%. The most notable AI procedures (Bayesian portrayal, k-NN, ANNs, SVMs) and of their suitability to the issue of spam.

* " An Approach for Spam Detection in YouTube Comments Based on Supervised Learning" by Amir Ali, Muhammad Zain Amin et al. (2016)

In this paper, they have evaluated several top-performance classification techniques for such purpose. The statistical analysis of results indicates that the Multilayer Perceptron and Support Vector Machine show good accuracy results of 94.89%.

* " A YouTube Spam Comments Detection Scheme Using Cascaded Ensemble Machine Learning Model” by Hayoung Oh et al. (2021)

# This paper proposes a technique to detect spam comments on YouTube, which have recently seen tremendous growth. Therefore, we examined related studies on YouTube spam comment screening and conducted classification experiments with six different machine learning techniques (Decision tree, Logistic regression, Bernoulli Naïve Bayes, Random Forest, Support vector machine with linear kernel, Support vector machine with Gaussian kernel) and two ensemble models (Ensemble with hard voting, Ensemble with soft voting) combining these techniques in the comment data from popular music videos - Psy, Katy Perry, LMFAO, Eminem and Shakira

# References

* <https://www.irjet.net/archives/V7/i4/IRJET-V7I488.pdf>
* H. Oh, "A YouTube Spam Comments Detection Scheme Using Cascaded Ensemble Machine Learning Model," in IEEE Access, vol. 9, pp. 144121-144128, 2021, doi: 10.1109/ACCESS.2021.3121508.
* Ali, Amir & Amin, Muhammad. (2016). An Approach for Spam Detection in YouTube Comments Based on Supervised Learning.