

# Spam Detection on Youtube Comments

BY Raka Arrayan



# BACKGROUND

Page 01

YouTube is the largest and most widely used video platform in the world, making it an invaluable marketing tool for companies. With billions of active users watching and interacting with content every day, YouTube offers a huge opportunity to reach a wide audience through advertising and promotions. However, the main challenge in utilizing this platform is ensuring that the company's ads and content are not disrupted by spam comments.

Spam comments, which can include irrelevant promotions, malicious links, or damaging criticism, often disrupt the user experience and can reduce the quality of the discussion beneath the video. Additionally, spam comments can create a negative image for the advertised brand, affecting the audience's perception of the company. Therefore, detecting and removing spam comments is crucial to maintaining the integrity of interactions on the platform. By using machine learning techniques and natural language processing, companies can effectively identify and filter spam, ensuring that their ads appear in a positive and relevant context.

# METADATA

Page 02

- **COMMENT\_ID:** Unique ID for each comment.
- **AUTHOR:** The name of the user or author of the comment.
- **DATE:** The date and time when the comment was posted. The format can vary, for example some are just the date and some are a complete timestamp.
- **CONTENT:** The content or text of the posted comment.
- **CLASS:** Label indicating whether the comment is spam (1) or non-spam (0)

# DATA LOADING

Page 03

	COMMENT_ID	AUTHOR	DATE	CONTENT	CLASS
0	z13lgffb5w3ddx1ul22qy1wxspy5cpkz504	dharma pal	2015-05-29T02:30:18.971000	Nice song	0
1	z123dbgb0mqjfxbtz22ucjc5jvzcv3ykj	Tiza Arellano	2015-05-29T00:14:48.748000	I love song	0
2	z12quxxp2vutflkxv04cihggzt2azl34pms0k	Priñcessé Ális Łovê Dømíñø Mâdiś™	2015-05-28T21:00:08.607000	I love song	0
3	z12icv3ysqvlwth2c23eddlykyqut5z1h	Eric Gonzalez	2015-05-28T20:47:12.193000	860,000,000 lets make it first female to reach...	0
4	z133stly3kete3tly22petvwdpmpghrlli	Analena López	2015-05-28T17:08:29.827000	shakira is best for worldcup	0
5	z12myn4rltf4ejddv23mwr3piuapcbl0r	jehoiada wellington	2015-05-28T17:06:37.288000	The best world cup song ever!!!!	0
6	z135vzqy1yrjhluew23kibopnrmqsplux	Kara Cuthbertson	2015-05-28T15:46:42.482000	I love	0

**CLASS**

0	196
1	174

Name: count, dtype: int64

Jumlah Train-Set : 259  
Jumlah Test-Set : 111

The image on the side shows the class numbers, where the class labeled 1 (spam) has 174 rows and the class labeled 0 (not spam) has 196 rows. It also shows the amount of data included in the training set, which is 70%, and the test set, which is 30% of the total data.

# VECTORIZATION

Page 05

```
# Inisiasi CountVectorizer  
  
from sklearn.feature_extraction.text import CountVectorizer  
  
vect = CountVectorizer()  
  
# Mempelajari kata apa saja yang ada pada `train_comment`  
vect.fit(train_comment)  
  
# Mengubah kata menjadi vektor berdasarkan frekuensi kemunculannya  
train_comment_mat = vect.transform(train_comment)  
test_comment_mat = vect.transform(test_comment)
```

Convert words to vectors based on their frequency of occurrence, and use the transform method to convert test data and test data to numeric based on their occurrence

# CLASSIFICATION

Page 06

```
from sklearn.neighbors import KNeighborsClassifier  
knn = KNeighborsClassifier(n_neighbors=5)  
knn.fit(train_comment_mat, train_class)  
  
from sklearn.metrics import accuracy_score  
  
hasil_prediksi = knn.predict(test_comment_mat)  
print('Akurasi : ', accuracy_score(test_class, hasil_prediksi))
```

Akurasi : 0.8108108108109

Performs classification with the K-Nearest Neighbors (KNN) technique by considering 5 nearest neighbors to determine the class of a data points and checking between The first parameter is the actual label, and the second parameter is the predicted label with an accuracy 81%

# NEW COMMENT TEST

Page 07

```
# Test komentar baru
komentar_baru = ["Your voice is very good and makes me inspired to always practice",
                  "Hey guys, subscribe to another channel that has a better voice and is also soothing"]

# Mengubah teks ke dalam bentuk vector
komentar_baru = vect.transform(komentar_baru)

# Memprediksi komentar baru
hasil_prediksi_komentar_baru = knn.predict(komentar_baru)
print('Hasil Prediksi : ', hasil_prediksi_komentar_baru)
```

Hasil Prediksi : [0 1]

In the picture above I did a test on a new comment, where I tried a test ‘Your voice is very good and makes me inspired to always practice’ and the machine guessed that this comment was not spam (0) then I also tested with ‘Hey guys, subscribe to another channel that has a better voice and is also soothing’ and the machine guessed that this comment was spam (1)

# CONCLUSION

Page 08

- Spam detection analysis of YouTube comments is an important step for content creators who want to maintain the quality of engagement on their channels. Comment spam can disrupt healthy discussions, reduce the quality of the viewer's experience, and damage hard-earned communities. By using machine learning techniques and natural language coding, creators can effectively identify and remove spam comments, ensuring that interactions on their channels remain positive and constructive. This not only helps in retaining a loyal audience but also increases creators' motivation to continue producing quality content.
- Spam detection analysis of YouTube comments is an important step for companies looking to advertise on the platform, as spam comments can harm the user experience and create a negative image for the brand. By using machine learning techniques and translating natural language, companies can effectively identify and filter spam, maintaining the quality of interactions on the channels chosen for advertising. This ensures that company advertising is displayed positively and in a relevant context, increasing audience trust and engagement, and optimizing the effectiveness of marketing campaigns.





Thank  
You!

