AD/CLICK FRAUD YOUTUBE VIDEOS

How to detect Videos promoting Ad/Click

Fraud

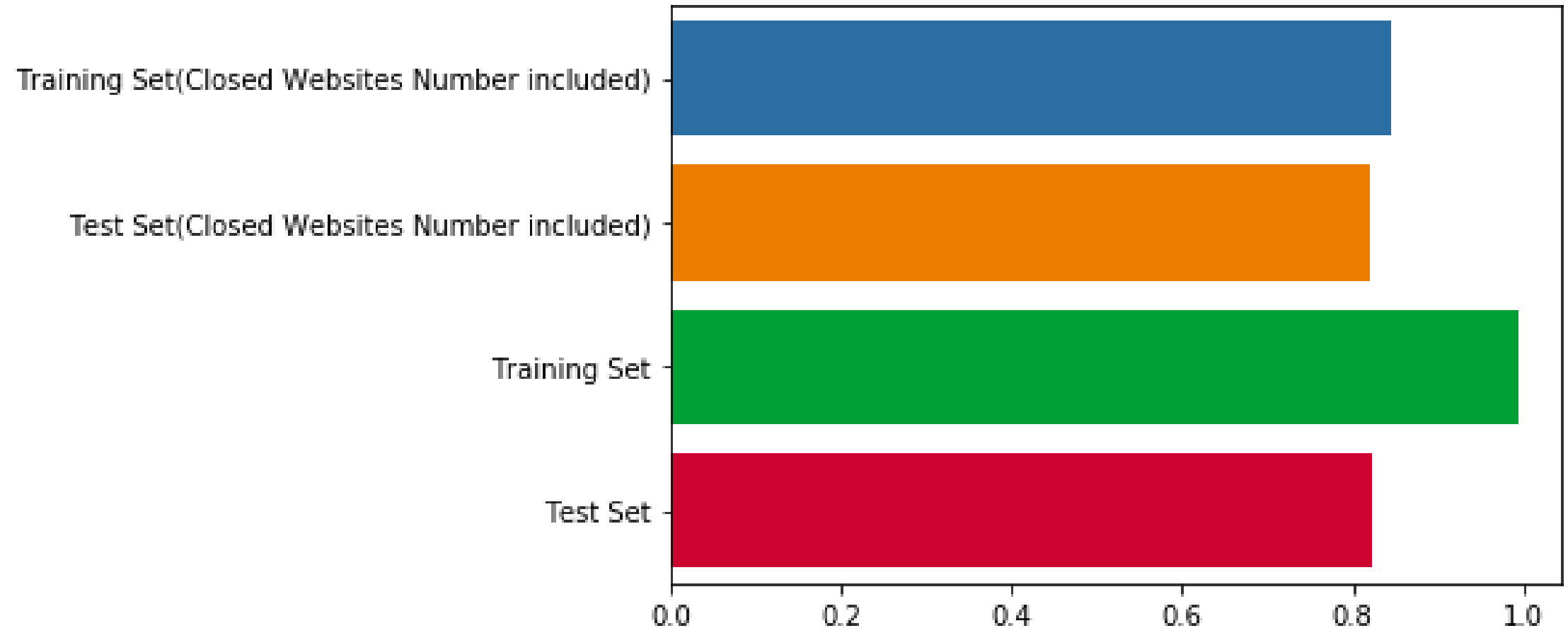# DATASET: YOUTUBE VIDEOS AD FRAUD DATASET

065

Dataset Has Following Features:

- Video_id
- Video_link
- Description
- Links(given in Description)
- Classification(as Fraudulent   or Benign)

Number of Records = **1065**

STATISTICAL INFERENCE/ MACHINE LEARNING MODELS USED

- Tf-idf Vectorizer used on Description
- SelectKBest
  - Present in sklearn.feature_selection
  - Used with chi-squared as Score Function
  - Used on Vectorized Features Obtained from Descriptions.
  - Select k=940 best features from these Vectorized Features
- train_test_split (from sklearn.feature selection) is used to split train and test data
- LogisticRegressionCV (from sklearn) is used to predict benign or fraudulent videos

INTERESTING FINDINGS/RESULTS

For Ad-Fraud research dataset must be up to date, and comprehensive because

- Fraudulent Videos are removed over time by Youtube

- Fraudulent Videos change their address or stop working in small amount of time.

Youtube Video Descriptions

- contain unique words specific to a specific video

- Problem of overfitting.

- Handled by using the k best features

Insights

- Ad Fraud Datasets must be comprehensive and Up to Date

- Selecting only the best features from Descriptions

  - Less Time for the model to train

  - Handles Overfitting

NEW THINGS LEARNT

- Tf-Idf Vectorizer

- SelectKBest

- Confidence Scores

THANK YOU