YouTube Fraud Detection

Muhammad Daniyal Dar, Rashna Kumar, Rohail Asim

1. Set up:

Our set up is such that we categorize videos into two sets:

* 1. Benign – Videos that do not promote fraudulent activities,
  2. Fraudulent – Videos that do promote fraudulent activities.

1. Data collection:

The dataset was collected using a web crawler. We were able to collect 1300 videos, using keywords namely:

* 1. Easy money,
  2. Make money online,
  3. Make money through clicking.

The first two resulted in mostly benign videos, which caused us to search for more precise results resulting in mostly fraudulent. Running a counter on 600 videos resulted in us finding 382 fraudulent videos in the 3rd search query.

1. Data categorization:

Although the videos were collected using a crawler and specific search terms, the categorization into benign and fraudulent videos had to be done manually, based on titles and video content. Initially, the videos had to be gone through manually and entirely to understand the cues that give away the fraud videos. After multiple videos, we were able to label videos mostly through titles exclusively, and had to go through the video content rarely.

1. Classification:

The next step is to come up with a mechanism to automate the process of classifying videos/uploaders into benign and fraud. This can be done through multiple methods, all possibly working concurrently:

* 1. Video collection:
     1. After finding a set of fraud videos, use YouTube’s rank based relevance algorithm to find related videos and then running the analysis algorithms on that video,
  2. Title Analysis:
     1. Videos that give away the content through titles directly through titles such as ‘Earn through clicking on ads/apps’.
     2. Titles that fall in the ambit of fraud, through ranking the text using LDA, further analysis might be required on them using some of the methods listed below in the following points,
  3. Description Analysis:
     1. Videos with descriptions giving away their potential content through both text and hyperlinks,
     2. Texts can be analyzed through LDA and LIWC, to understand the categories in which the texts belong to,
     3. Hyperlinks can be visited and their texts on the websites can be analyzed in the same way as 4)b.iii.,
     4. Hyperlinks to applications can be analyzed through their descriptions in the same manner,
  4. Comment Analysis:
     1. LDA/LIWC to rank texts, understand contexts and use them to determine the potential category the video might belong to,
  5. User Analysis:
     1. Uploaders are characterized by all the tags they have used on uploaded videos,
     2. The type of content can therefore be identified and the uploader can therefore be ranked/matched based on their upload profiles,
     3. Analysis on the subscriptions/favorites of identified fraudulent uploaders can help track down further potential fraudulent uploaders and users through snowball sampling.