**Content Moderation**

* This service is used to identify malicious content from the documents.
* Supported documents are Pdf, Word, Excel, PPT and Images.
* Backend code deployed on "10.104.124.26" On Premise GPU server.

**Files Used:**

1. embedding.py: [content\_moderation/embeddings.py]

"oletools" - python library is used for identify embedding in the documents. If document contains .sh or .exe file then it is malicious content.

1. image\_moderation.py : [content\_moderation/classification/image\_moderation.py]

**[On branch azure\_integration]**

**Implemented Solution:**

Azure image Moderation API is used for Image moderation as below alternative solution was time consuming.

**[On branch master ]**

**Alternative solution for skipping Azure Image Moderation API:**

To identify image moderation, "mobilenet" model is used. This model classify the image into 5 category-[drawings, hentai, neutral, porn, sexy]. If the classified image threshold value is greater than define threshold level of classified categories, then the image is malicious.

**API:**

* app.py – This files contains the API of content moderation service.

**Post a file for scanning :**

* URL: http://10.104.124.26:8001/content\_moderation
* Method: POST
* Request:
* file : file stream
* Response: {malicious\_content: boolean}

**Repository Links:**

* <https://devstack.vwgroup.com/bitbucket/scm/~gxn1f5r/content_moderation.git>