**Objective:**

The objective of the project is to find out whether student assignment is similar or not. As most of the students are found copying the assignment from each other. So, this application helps to find out many similarities are there between each of the uploaded assignments for example if there are 4 files uploaded it will compare (1,2) (1,3) (1,4) (2,3) (3,4) (2,4). It compares all the upload files with one another and prints the result in the console.

**File and Image supported:**

It can check the similarity between the image whether it is a pdf or JPEG. However, it compares the same type of file image with image and pdf with the pdf. Furthermore, it can check similarities between the pdf but the pdf should consist of selectable text but can’t however it can contain images, numbers, special characters, or Greek symbols.

**Input and output:**

You can provide input on the GUI interface that opens when you run the project and the output will be displayed on the console.

**Before you run the project make sure the following things:**

1. Python-tesseract is an optical character recognition (OCR) tool for python. That is, it will recognize and “read” the text embedded in images. ##If you don't have tesseract executable in your PATH, include the following: # pytesseract.pytesseract.tesseract\_cmd = r'<full\_path\_to\_your\_tesseract\_executable>' # Example tesseract\_cmd = r'C:\Program Files (x86)\Tesseract-OCR\tesseract'

2. Install all the modules used in the program.

such as tkinter, PyPDF2, pytesseract and so on you can find the module name in the program itself.