پروژه شناسایی ویروس:

هدف از این پروژه پیدا کردن فایل در صورت ویروسی بودن بود که برای این کار ابتدا بعد از بررسی چند نمونه فایل ویروس متوجه میشویم که فایل های ویروس substring های مشترک دارند.
با بررسی فایل ها متوجه شدم که تعداد بسیار زیاد کاراکتر های این فایل ها،کاراکتر و string های پرت هستند و می توان ان ها را به عنوان داده پرت در نظر گرفت.
برای این کار من تمامی کد ascii این کاراکتر ها مانند 0،@,# و ... پیدا کردم و ان ها را در لیست ذخیره کردم.
با بررسی بیشتر متوجه میشویم که فایل های هر فولدر برای مثال 50 تای متوالی شباهت های زیادی دارند.
برای همین برای پیدا کردن پترن ها هر 15 تا فایل متوالی را در نظر گرفتم و substring مشترکشان که یعنی string ای که از حروف انگلیسی است پیدا کردم،یعنی 15 فایل اول سپس فایل 35 تا 50 و به همین ترتیب فایل ها را انتخاب کردم زیرا اشتراک های 15 تای اول تقریبا همان اشتراک ها 35 تای اول است.

در تابع `search_in_folder` کاری که می کند این است که ابتدا تمام 20 فولدر را باز میکند و در فایل های هرکدام پیمایش میکند.
سپس به ازای هر فولدر یک فایل txt میسازم و تمامی پترن هایی که از هر فولدر پیدا شده را در ان قرار دادم.

```
def search_in_folder():
    for i in range(20):
        Folder = f"C:\\Users\\USER\\Desktop\\AD_project\\Released\\Train\\Malware Sample\\{i}"
        for f in range(len(os.listdir(Folder))):
            while(f<(len(os.listdir(Folder)))):
                file2search = os.path.join(Folder,os.listdir(Folder)[f])
```

برای اضافه کردن substring شرط هایی مثل حداقل 5 و حداکثر 25 را اعمال کردم.

سپس تابع FindCommon را صدا میزنیم.این تابع در هر فایل ساخته شده به ازای هر فولدر در مرحله قبل پیمایش میکند و به ازای هر line تمام فایل را بررسی میکند و تعداد دفعات تکرار این line را محاسبه می کند و اگر بیش از تعدادی برای مثال 50 تا بود به ListOfPatterns که لیستی از پترن هاست اضافه میکند.

```
def FindCommon():
    Folder = "C:\Users\USER\Desktop\AD_project\PatternOfFolders"
    for f in range(len(os.listdir(Folder))):
            while(f<(len(os.listdir(Folder)))):
                file2search = os.path.join(Folder,os.listdir(Folder)[f])
    with open(file2search) as f:
        data = f.readlines()
    for line in data:
        count = 0
        for checkline in data:
            if (line == checkline):
                count+=1
        if (count>=50):
            if line not in ListOfPattern:
                ListOfPattern.append(line)
```

سپس تابع DeletePattern را صدا میزنیم.این تابع در تمام فایل های بی خطر پیمایش میکند و اگر عضوی از ListOfPatterns در آن باشد آن عضو را از این لیست حذف میکند.

در نهایت تابع CheckFile تابعی است که فایلی از ورودی میگیرد و تمام لیست ListOfPatterns را در ان بررسی میکند.اگر substring ای پیدا شد یعنی ان فایل را ویروس تشخیص داده است.

صباکیانوش