Detection of Malicious URLs using Machine Learning Techniques

MAIN PROJECT

Dept Of MCA ,MES COLLEGE OF ENGINEERING, KUTTIPPURAM

07 MAY 2021

SCRUM TEAM

SCRUM MASTER

SOUDHA A M (LMES18MCA11056)

PRODUCT OWNER

MR. NOWSHAD C V

ASSOCIATE PROFESSOR

DEPARTMENT OF COMPUTER APPLICATIONS

MES COLLEGE OF ENGINEERING, KUTTIPPURAM

Table of Contents

- Introduction
- Modules
- Architecture
- User Story
- Project Plan
- Product Backlog
- Sprint1
- Developing Environment
- Git Details

Detection of Malicious URLs using Machine Learning Techniques

The primitive usage of URL (Uniform Resource Locator) is to use as a Web Address. However, some URLs can also be used to host unsolicited content that can potentially result in cyber attacks. These URLs are called malicious URLs. The inability of the end user system to detect and remove the malicious URLs can put the legitimate user in vulnerable condition. Furthermore, usage of malicious URLs may lead to illegitimate access to the user data by adversary. The main motive for malicious URL detection is that they provide an attack surface to the adversary

It is vital to counter these activities via some new methodology.There have been many filtering mechanisms to detect the malicious URLs. Some of them are Black-Listing, Heuristic Classification etc. These traditional mechanisms rely on keyword matching and URL syntax matching. Therefore, these conventional mechanisms cannot effectively deal with the ever evolving technologies and webaccess techniques.

Detection of Malicious URLs using Machine Learning Techniques

Furthermore, these approaches also fall short in detecting the modern URLs such as short URLs, dark web URLs. We propose a novel classification method to address the challenges faced by the traditional mechanisms in malicious URL detection. The proposed classification model is built on sophisticated machine learning methods that not only takes care about the syntactical nature of the URL, but also the semantic and lexical meaning of these dynamically changing URLs.

MODULES

WWW/URL DATABASE

WWW or URL database is a component from where number of URLs are fetched. These URLs from multiple websites are collected using the web crawler and are stored in the URL database.

FEATURE EXTRACTION

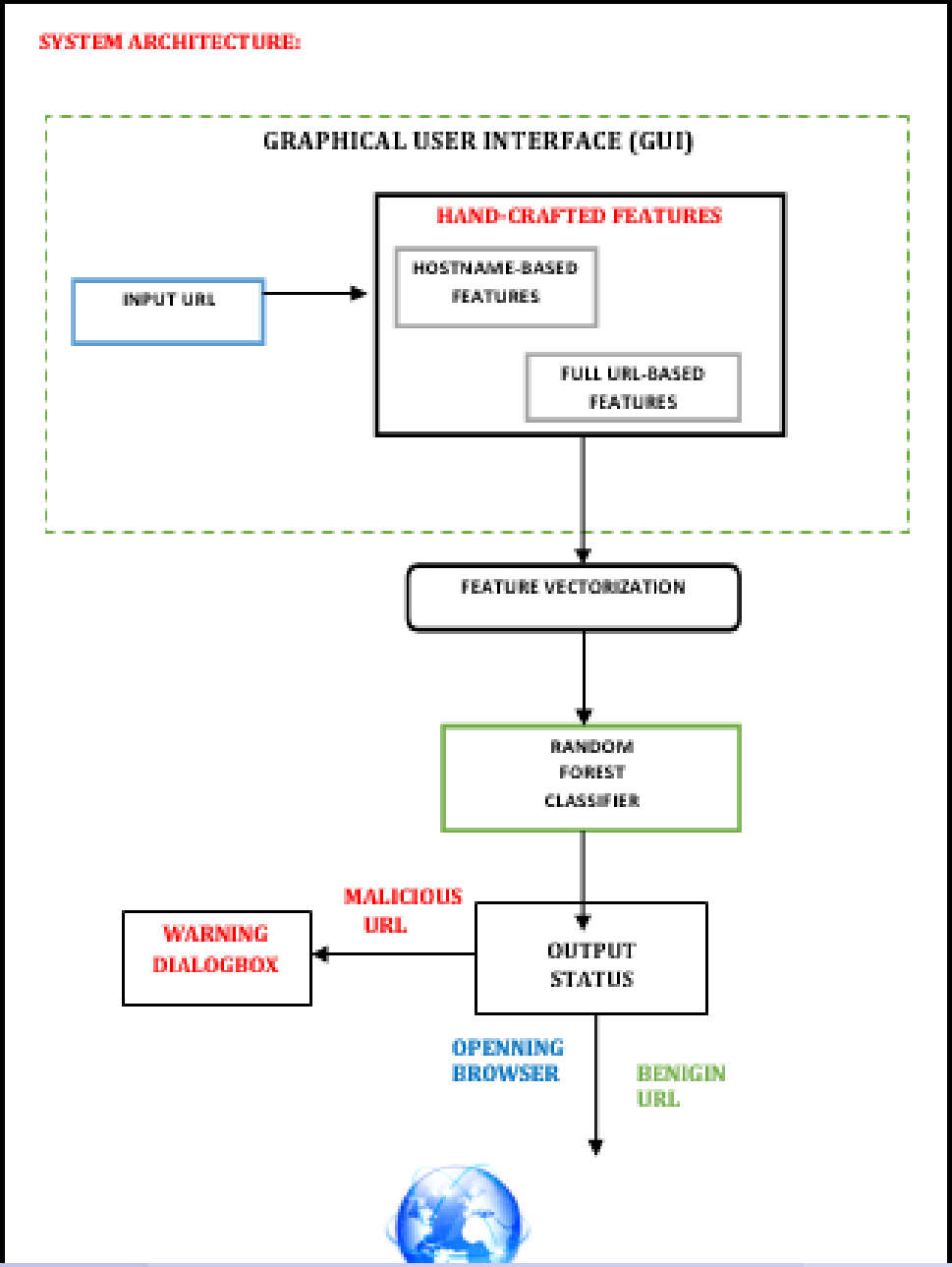
This component is used to extract the features from the URL. If the URL already exists in the blacklist then it is qualified as a malicious. In this component it will classify the URL also on the basis of lexical features.

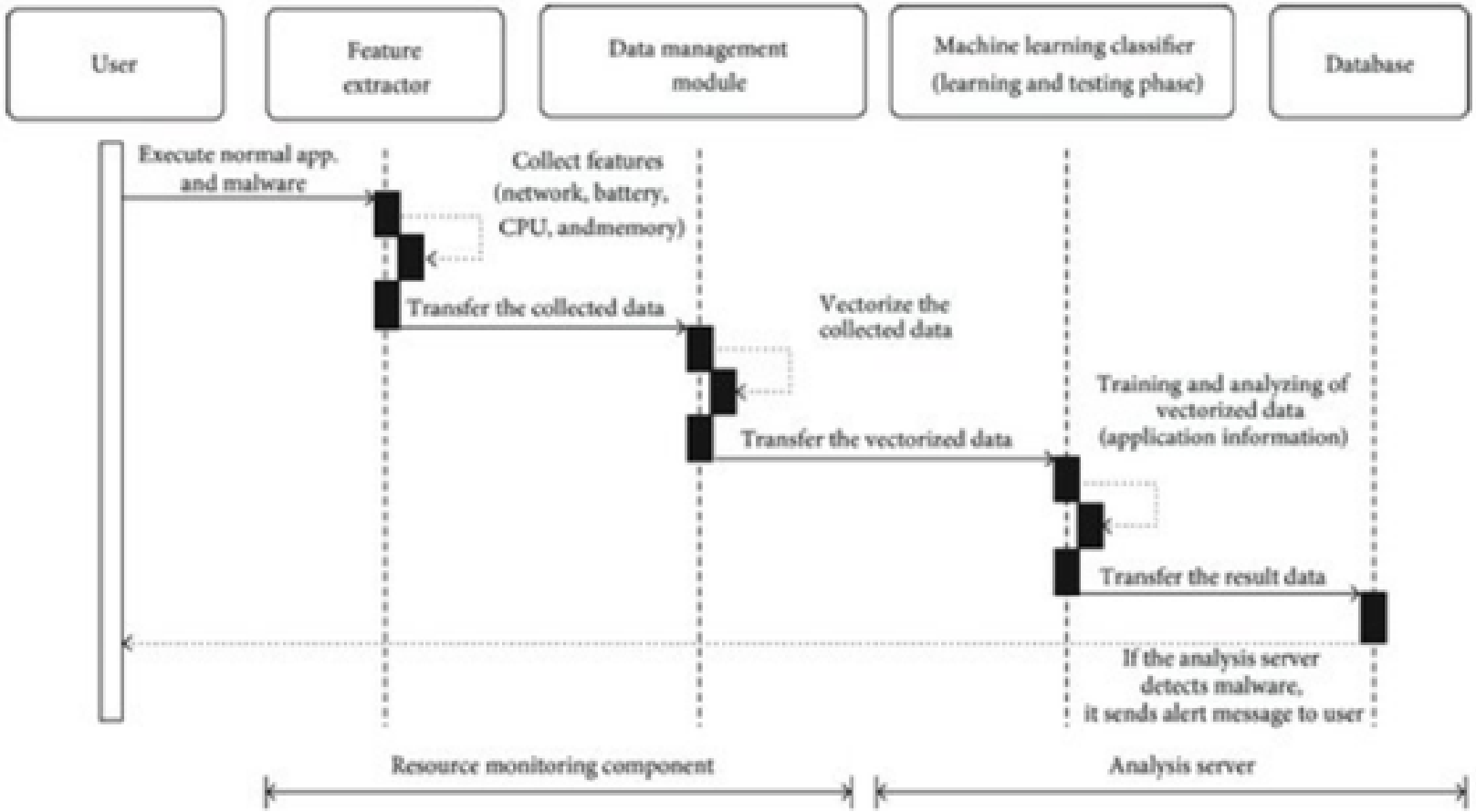
MODULES

MACHINE LEARNING CLASSIFIER

This component is used to classify the URL whether it is a malicious or benign. It is done on the basis of features collected by the previous component. Previous component's result will serve as an input to this component. After the classification, the URLs will be classified as whether it is malicious or benign.

ARCHITECTURE




User
Feature extractor
Data management module
Machine learning classifier (learning and testing phase)
Database
Execute normal app. and malware
Collect features (network, battery, CPU, andmemory)
Transfer the collected data
Vectorize the collected data
Transfer the vectorized data
Training and analyzing of vectorized data (application information)
Transfer the result data
If the analysis server detects malware, it sends alert message to user
Resource monitoring component
Analysis server

USER STORY

ID	As a <type of user>	I want to <perform some task>	So that I can <achieve some goal>
1	User	Input URL	Enter URL to see if it is malicious or benign
2	User	Submit button	To identify whether URL is benign or malicious. To prevent the browser from opening if it is benign. If it is malicious it will show a warning message and prevent it open as well.

PRODUCT BACKLOG

ID	NAME	PRIORITY	ESTIMATE DAYS
1	URL DATASET	1	3
2	FEATURE EXTRACTION	2	8
3	MACHINE LEARNING CLASSIFIER	2	10
4	ANALYSE URL	2	10
5	OPEN BROWSER	2	4
6	WARNING DIALOGBOX	3	4

PROJECT PLAN

User Story ID	Task Name	Start Date	End Date	Days	Status
1	Sprint 1	28/03/2021	3/04/2021	8	Completed
	Form designing	28/03/2021	30/04/2021	2	Completed
	Coding	31/03/2021	5/04/2021	5	Completed
	Testing	6/04/2021	07/04/2021	1	Completed
2,3	Sprint 2	08/04/2021	26/04/2021	18	Completed
	Data collection	08/04/2021	15/04/2021	7	Completed
	Coding	15/04/2021	25/04/2021	10	Completed
	Testing	25/04/2021	26/04/2021	1	Completed
4,5	Sprint 3	04/05/2021	20/04/2021	15	In process
	Coding	04/05/2021	18/05/2021	14	In process
	Testing	19/05/2021	20/05/2021	1	In process
6	Sprint 4	25/05/2021	07/06/2021	14	pending
	Coding	25/05/2021	03/06/2021	10	pending
	Testing	04/06/2021	07/06/2021	4	pending

SPRINT1

Backlog item	Completion Date	Estimate in hours	Day1 28/03	Day2 29/03	Day3 30/03	Day4 31/03	Day5 01/03	Day6 02/03	Day7 03/03	Day8 04/03	Day9 05/03	Day10 06/03	Day11 07/03
User story	#1	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)	(H)
Form designing	30/04/2021	8	4	2	2	0	0	0	0	0	0	0	0
coding	05/04/2021	16	0	0	0	6	2	2	2	2	2	0	0
Testing	07/04/2021	4	0	0	0	0	0	0	0	0	0	2	2
Total		28	4	2	2	6	2	2	2	2	2	2	2

DEVELOPING ENVIRONMENT

FRONT END	:	PYTHON
IDE	:	PYTHON IDLE
BACK END	:	URL DATASET

GIT DETAILS

SoudhaAM / Detection-of-Malicious-URLs

Unwatch 1 | Star 0 | Fork 0

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

main | 1 branch | 0 tags | Go to file | Add file | Code

SoudhaAM Feature extraction added — cdb9c78 1 minute ago — 6 commits

Feature_extraction.py	Feature extraction added	1 minute ago
OIP.png	gui added	3 days ago
README.md	Initial commit	4 days ago
gui.py	gui added	3 days ago
url_features.csv	url dataset	3 days ago

README.md

Detection-of-Malicious-URLs

Detection of Malicious URLs using Machine Learning Techniques

About

Detection of Malicious URLs using Machine Learning Techniques

Readme

Releases

No releases published
Create a new release

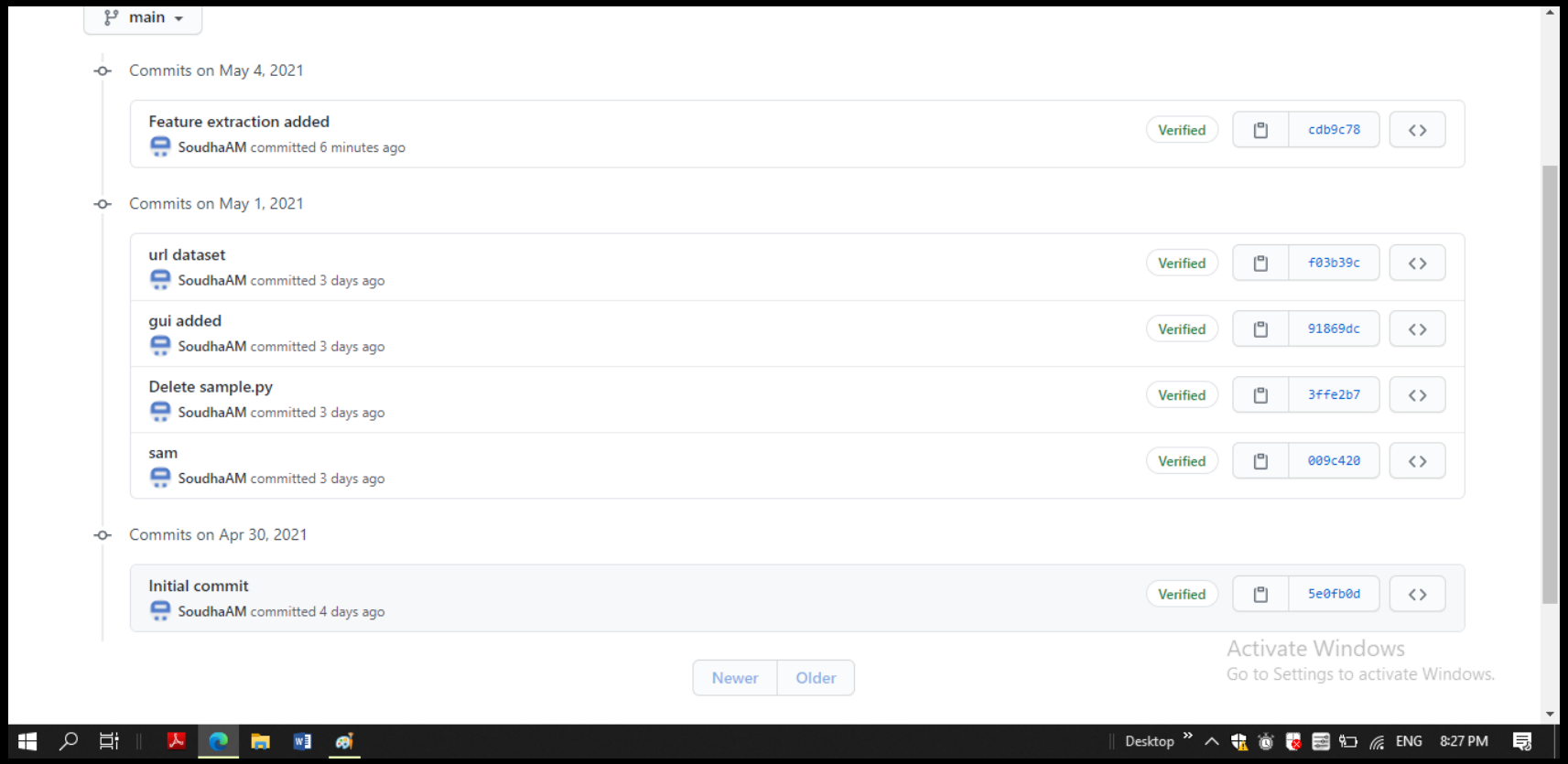
Packages

No packages published
Publish your first package

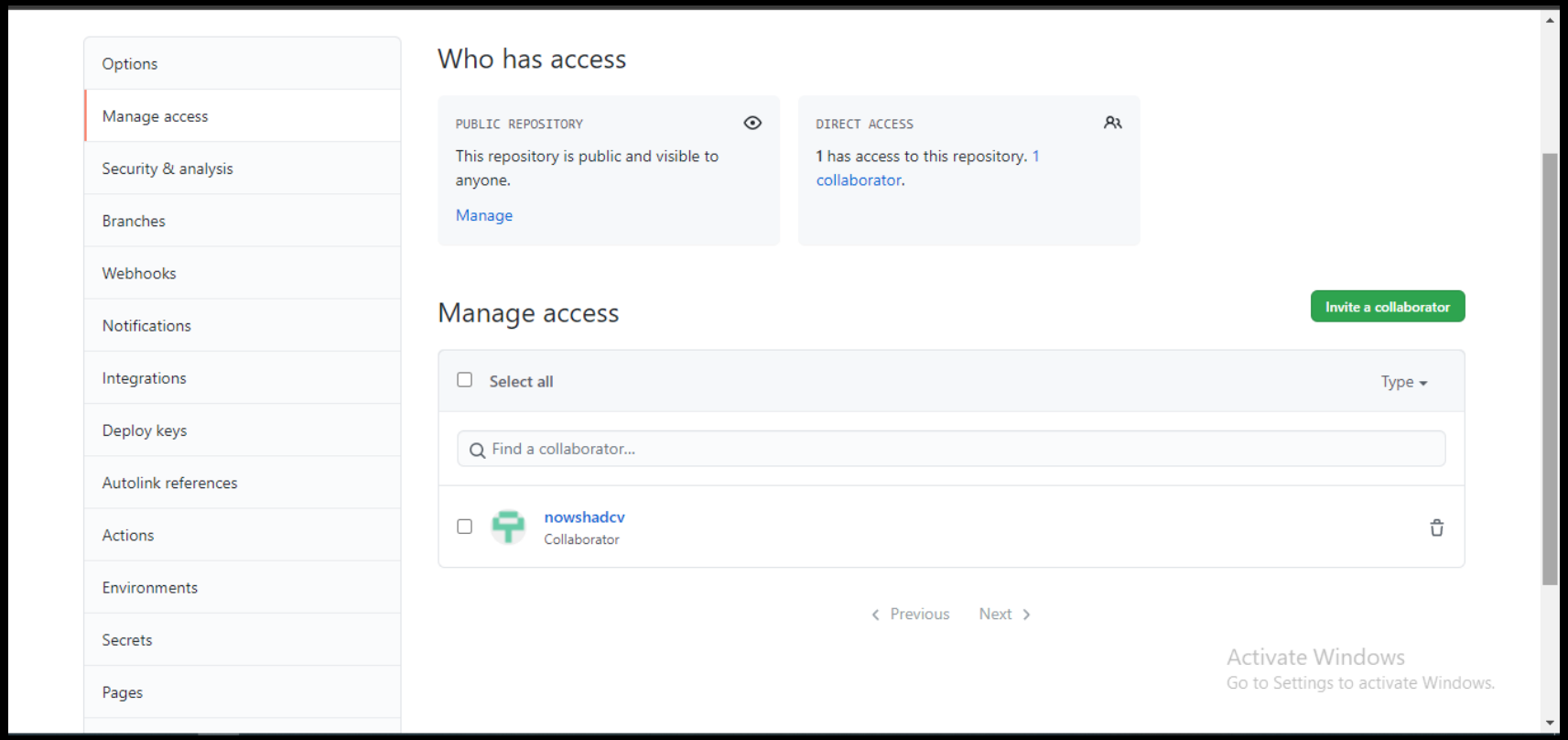
Languages

Python 100.0%

GIT DETAILS


main
Commits on May 4, 2021
Feature extraction added
SoudhaAM committed 6 minutes ago
Verified
cdb9c78
Commits on May 1, 2021
url dataset
SoudhaAM committed 3 days ago
Verified
f03b39c
gui added
SoudhaAM committed 3 days ago
Verified
91869dc
Delete sample.py
SoudhaAM committed 3 days ago
Verified
3ffe2b7
sam
SoudhaAM committed 3 days ago
Verified
009c420
Commits on Apr 30, 2021
Initial commit
SoudhaAM committed 4 days ago
Verified
5e0fb0d
Activate Windows
Go to Settings to activate Windows.
Newer
Older
Desktop
ENG
8:27 PM

GIT DETAILS


Options
Manage access
Security & analysis
Branches
Webhooks
Notifications
Integrations
Deploy keys
Autolink references
Actions
Environments
Secrets
Pages
Who has access
PUBLIC REPOSITORY
This repository is public and visible to anyone.
Manage
DIRECT ACCESS
1 has access to this repository. 1 collaborator.
Manage access
Invite a collaborator
Select all
Type
Find a collaborator...
nowshadcv
Collaborator
Previous
Next
Activate Windows
Go to Settings to activate Windows.

THANK YOU