Universidade de Aveiro

Departamento de Eletronica, Telecomunicações e Informática

PEI Team 07

ProtectMe : Elaboration

Contents

Introduction

In recent years, there has been an enormous growth in social media platform usage and influence on the masses.

This growth can mostly be attributed to the rise in usage of mobile computing devices and the now easier access to the internet (mostly through the usage of wireless technologies).

Although the usage of social media has come to improve the life of the common person, be it through facilitating the spread of news or offering an easier way to communicate with one's peers. But with this spur in growth of people who use social media platforms in order to improve their lifestyle, there are also people who abuse these platforms in order to spread malicious content and fulfil a given agenda.

1. Requirements Elicitation

The specification process can be divided in two stages: firstly, an assessment of the main goals and objectives we expect to fulfil, in order to define the scope of this project; and secondly, an evaluation of the current state-of-the-art (projects and academic papers on this subject and related to the integral parts of this project) so as to understand what has and hasn't yet been done in this context.

1.1   Context

This subsection presents a description on how the system is expected to be used. The user (someone who wants to check if a given social media post contains any malicious information) opens the application's web page and inserts the URL to the social media post in the application's text box. This will be the only way to verify a new social media post.
After inserting the URL in the text box, the user should click on the Verify button and wait for a moment, for the request to be processed. After waiting the required time, the user will be taken to a page with all the data extracted from the post and an evaluation of whether or not the post may contain malicious information. This will be the main usage of the application.

1.2   Objectives and Purpose

The ProtectMe! project aims to provide the means to monitor/verify the spread of malicious content in a social media context through the usage of Natural Language Processing(NLP) algorithms , Media Content Extraction Tools and Machine Learning Software. As such, by the end of the project, we expect to have completed the following objectives:

- Gather a list of functional and non-functional requirements

- Analyze the current State of the Art

- Produce a content extraction pipeline

- Find a Viable Data Set in order to train the machine learning model

- Produce and train a machine learning model

- Produce a web interface to analyze the results of the machine learning model

1.3 State of the Art

Although the spread of malicious content on social media platforms has always been a very serious problem in this type of web services, there hardly ever was an attempt to monitor/detect the spread of this type of content making use of Natural Language Processors (NLP) and Media Content Extraction Algorythms (mostly because only recently have these technologies become ready for generic usage). As such, most works done in this area have mostly been focused on detecting spam attacks and offensive content spread (mostly detecting the spread of gore/pornographic content regardless of the context). The best example we could find that's related to this kind of work was done by Johnny Wales (Wales Technology LLC) in November, 2019. ¡Insert Link to Article¿ In this article, Wales explains the process he went through in order to create and train a Machine Learning Model that would classify news articles as either normal or fake news based on the way the article itself was written. Of course this article by itself has no immediate connection to the work we are trying to develop in this project. However, the usage of NLP in order to categorize text in an article solely based on the way it was written is part of the aim of our project.

2. Users

This application only foresees one kind of user: a simple user who wants to check if a social media post contains malicious content or not.
From this original type of user, we can sub-categorize it into a registered or unregistered user,
A unregistered user would be someone who wants to anonymously check a single post, as a single time event whereas a registered user would be someone who would be using this service with some manner of frequency and would like to keep track of the post they searched and the results they got,

2.1 Use Cases

In this section, we will be defining what the use cases for each type of user are, as well as clarifying the key aspects of each user.

2.1.1 Common User

The common user is a user who uses the system in an anonymous manner, who doesn't desire to have a track record of his searches. As such, it's only supposed to query a given social media post and analyze the results obtained.

2.1.2 Registered User

A registered user is a user who uses the system in some sort of regular manner, and has the need to keep a track record of his queries. As such, this user is able to log into the application (as well as log out); check his search history and, like the common user, query a social media post and analyze the query's results.

3. Non Functional Requirements

In this section, we will the approach the application's non functional requirements, along with a priority evaluation and a short description of the decision for the priority attributed to a given requirement.

3.1 Usability

Due to the nature of this application, it must be easy to use and easy to learn. Although there are no functionalities that require complex knowledge about the application, it still deals with data and statistics when it comes to show the results of a given query to the user. As such, it is on our best interest to make the application easy to learn and to give relevant feedback to the user.

Priority: High

3.2 Reliability

As the system's purpose is to predict if a given social media post contains malicious content, the reliability of the given results would be one of the most important aspects of this project. However, we won't be able to give 100% accurate results. Far from it. As such, we want to give feedback to the user on how reliable a given result is.

Priority: Average

3.3 Security

Although our application doesn't really deal with users' personal data, there won't be an enormous focus on the service's security.

Priority: Low

3.4   Interoperability

There isn't any interoperability planned for this project since it is at the moment regarded as a standalone web application.

Priority: None

3.5   Portability

Due to the nature of a web application, the application must be accessible by different browsers and should operate in the same way for each one, be it desktop or mobile mode.

Priority: High