Associate Professor Catalin BOJA, PhD
Email catalin.boja@ie.ase.ro
The Bucharest University of Economic Studies
Lecturer Alin ZAMFIROIU, PhD
Email alin.zamfiroiu@csie.ase.ro
The Bucharest University of Economic Studies
Assistant Madalina ZURINI, PhD
Emailmadalina.zurini@csie.ase.ro
The Bucharest University of Economic Studies
Assistant Bogdan IANCU, PhD
Emailbogdan.iancu@ie.ase.ro
The Bucharest University of Economic Studies

USER BEHAVIOUR PROFILING IN SOCIAL MEDIA APPLICATIONS

Abstract: *Given the increase of sensitive data stored within mobile devices and social media applications, the need for creating a user behaviour profile is in high demand for preventing security breaches, impersonation or unauthorized access to resources. The behaviour of a user is defined by the aggregation of different patterns that are obtained while constantly using a software service on a mobile device or a computer. The scope of this study is to identify a core group of characteristics that can be further used in profiling a user based on his behaviour. For that, a survey consisting of 20 questions was conducted having a set of 356 responders. The obtained data were pre-processed and used as input for supervised and unsupervised classification techniques.*

Two applications, one web-based and one mobile, were implemented in order to expand the prior selection of characteristics and to verify the obtained results. They were used to measure interaction events and ways of using social media on both versions of the applications. Taking into account these different behaviour based characteristics, we defined different groups of users that are uniquely identifiable by limited sets of characteristics.

Keywords: *social media applications, user profiling, user behaviour, supervised and unsupervised classification*.

JEL Classification: C63, C83, C88

1. Introduction

Social network platforms usage nowadays has skyrocketed with the increasing number of smartphones connected to the internet. Even though social networks

DOI: 10.24818/18423264/53.1.19.02

existed since 1997 (Boyd & Ellison, 2007), it took more 15 years for one platform to reach 1 billion users worldwide (Zuckerberg, 2012).That one is Facebook, the most used social network with more than 64% of the market share in the UK (Statista, 2018)and with similar percentages worldwide. Unfortunately, its popularity attracted also entities with not-so-good intentions. The problem of fake news is still a debatable one together with the one related to the fake accounts (Hunt & Gentzkow, 2017). Another major problem is one of the data leaks, caused either by bugs in the platform (Heaven, 2018), either by phishing attempts which convince the users into exposing their account credentials.

In this paper, we propose a behaviour analysis in order to identify the key characteristics of Facebook social platform usage (either on mobile phones or on PCs). Once the characteristics that have the biggest influence on user behaviour profile are identified, we can create especially designed applications that run the social network in “protected mode”. By measuring these characteristics in real time and comparing them with the ones measured before, the software can identify account theft. We also analysed the identified characteristics from a statistical point of view and we applied different classification techniques in order to find patterns in the users’ behaviour, patterns that can trigger the alarm once they are broken. By doing all of these we can add a security layer on top of the standard ones so that the users can browse safely their social media accounts. The same behaviour profile can be also used to distinguish between human users and machine controlled actions, like bots and other types of software that can impersonate user behaviours. Also, determining a human behaviour model and being able to measure it, has advantages in other fields like Information Security(Sasse, Brostoff, & Weirich, 2001), (Kayacik, Just, Baillie, Aspinall, & Micallef, 2014)or Machine Learning(Webb, Pazzani, & Billsus, 2001).

2. User interaction with social media platforms

In order to analyse the interaction of users with social applications, a questionnaire was conducted for interviewing an adequate number of people about how they are using the social media platform Facebook.

In the first part of the questionnaire, information about the respondents was collected, such as age, sex, location, the university where they are studying, and the year of study (for those who are still studying).In the second module, questions were added to find out how much time the users spend on one of the applications made available by the Facebook platform: mobile Facebook application, Messenger for mobile communication or website via browser. Also, within this module, we tracked the way a user interacts with the mouse or the keyboard in the applications made available by the Facebook social media platform was tracked. In the third questionnaire module, the focus was on how users interact with Facebook applications. The final scope was to obtain a large number of behavioural features related to user interactions with social media platforms. The questions were constructed so the respondent answers will reveal:

- what is the main platform they use to access their social networks accounts, mobile version versus Web one
- what are the human factors that defines the profile, in a machine versus human analysis; from this perspective we focused the questions on user actions that are bound to the fact that humans are prone to do mistakes when they interact with a device or software applications; different kind of mistakes are related to not paying attention or being distracted when you type or select a control or simply to not knowing how to use it; therefore users are generating multiple actions for correcting or changing their content before publishing it and we wanted to check how this will affect their behaviour.

A number of 356 people were questioned through this questionnaire. The analysis of the results showed that 46% of users spend 1-2 hours daily on the Facebook social media platform through their mobile apps, as it is presented in Figure 1.

In the web application, 79% spend daily less than an hour on the Facebook social media platform. From these two measurements, it can be concluded that users are preferring the mobile devices to interact with the social network rather than the web application.

To the question "How do you scroll in Facebook applications", 64 people responded that they make a scroll with the mouse, 18 people with the keyboard, and 274 people with their finger on the touch screen. This was somehow expected because many respondents are using the mobile application on Facebook or Facebook Messenger.

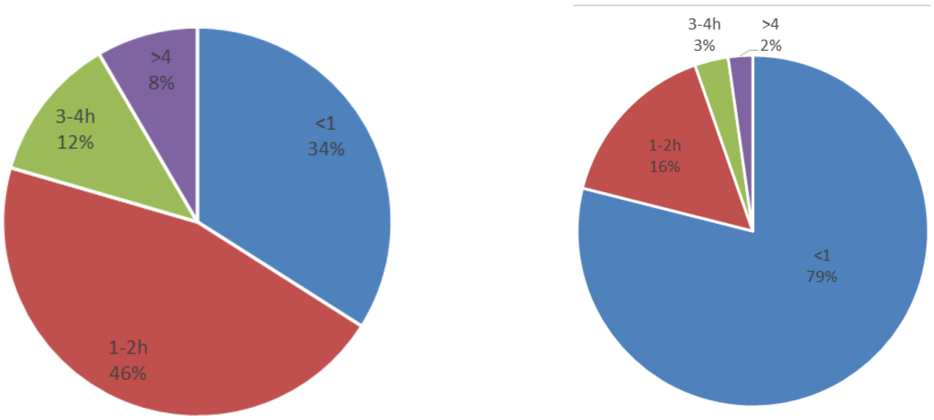


Figure 1.Time spent on Facebook (mobile app - left and Web app - right)

Regarding the correction of typos, 22 people said they moved the cursor to the beginning of the wrong word and deleted it using the DELETE key, 115 said they moved the cursor to the end of the wrong word and erased it with the BACKSPACE key, 182 said they were selecting the wrong word or the wrong letters and rewriting that selected part, while 37 people said they delete everything they wrote after mistake until they got to that mistake and then rewritten everything. These results are shown in Figure 2.

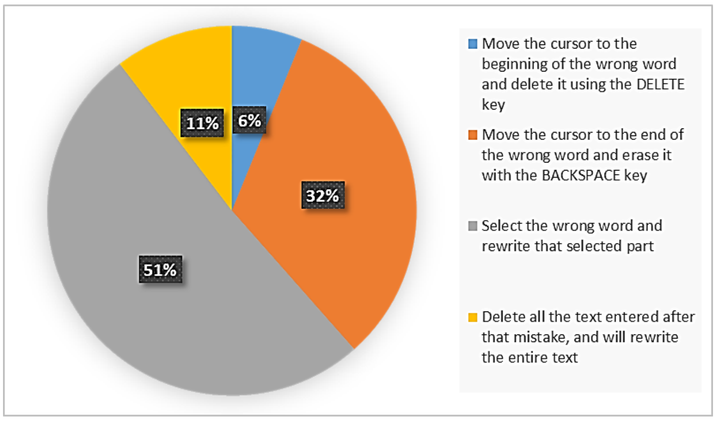


Figure 2. Diagram of responses on methods used for deleting a typo mistake

Moving the cursor is done in most of the cases by touching the screen of the mobile device. The last question related to how the users are holding the mobile device was answered in the following way: 108 people answered that, when they read a text on the mobile device, they are using it in landscape mode, and 248 people in portrait mode. To make a scroll on Facebook 41 people used the device in landscape mode, and 315 in portrait mode. To write texts in Facebook applications, 44 people use their device in landscape mode and 312 use it in portrait mode. These results are shown in Figure 3.

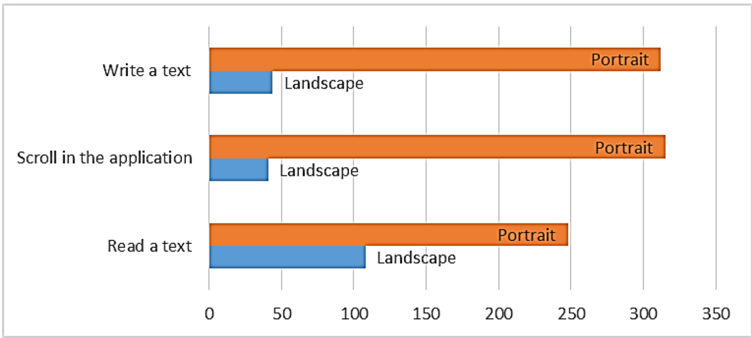


Figure3. Results of the question on "Ways of using the mobile device"

We conclude that, in most of the cases, the respondents are using mobile devices in portrait mode. This result is somehow explained by the fact users want to see more information on the same screen and to scroll it from top to bottom.

Based on the results obtained from the analysis of the user interaction questionnaire, we extracted the user interaction characteristics with social media applications that can be used to define the behaviour model.

3. Characteristics of the social media applications users' behaviour

The analysis of the questionnaire responses shows that most respondents are using mobile devices to interact with the Facebook social networking platform. This was excepted as other studies (Enge, 2018) conducted on how users are accessing Internet resources have been highlighted this.

The interaction characteristics that can be considered for determining the behaviour of users in social media applications are:

- **scroll mode**; scrolling can be done with the mouse or the keyboard for web applications or finger dragging on mobile devices; in the case of web applications, it can be determined which mode is preferred by the user; the most used way for each user can be saved in his profile; for mobile applications, because only the finger is used to drag, it is possible to determine the area on the display used to perform this action; this characteristic will depend heavily on the size of the device; but a user usually uses the same finger to scroll and, depending on how the phone is used, the scrolling is made on the left side of the screen, on the right side or centre of the screen of the mobile device. This feature is also influenced by the content of the page that the user is scrolling through because, if the page contains buttons on one side, the user will try to avoid that area of the page, and will scroll throughout the other side of the screen.
- **the importance given to the typing mistakes**; to measure this feature, the text entered by the user must be checked if it contains or not typing errors; in this way, two types of users can be distinguished:
    - users who ignore the typing mistakes;
    - users who correct the entered text each time.

  For the second category, those who correct the input text, the following characteristics can be measured.
- **correction mode of the typing mistakes**; some users when they notice that they have a wrong word will delete all the entered text after that mistake, and will rewrite the text; other users will move the cursor very close to the mistake and will erase only the wrong text, then they will enter the correct one; the value of this indicator can be determined by the fact that the user moves the cursor or not, and after the move one deletes the text; this feature applies to web applications and mobile applications;
- **deletion mode of the text**; this characteristic is specific to web applications on Windows because with a standard keyboard, the text can be deleted with the DELETE key or the BACKSPACE key; so there are two types of users:
    - users who move the cursor to the end of the wrong word and delete using the BACKSPACE key;

  - users who move the cursor to the beginning of the wrong word and delete using the DELETE key.

  this characteristic is not specific to mobile applications because for these applications the text can be deleted only with the BACKSPACE key;
- **the way of using the mobile device to read texts**; many users are using the mobile device only in PORTRAIT mode but, in some specific moments, when they have a lot to read, they use it in LANDSCAPE mode; this feature can be a component of the user's interaction profile with the mobile application;
- **the way of using the mobile device to write text**;to have a larger virtual keyboard, some users are using the mobile device in LANDSCAPE mode; this feature is also important for achieving the user profile in the interaction with mobile applications;
- **the way of using the mobile device to scroll newsfeeds in social media applications**; most users use the device in PORTRAIT mode for this action but, if a user frequently uses the device in LANDSCAPE mode, this can be considered as an obvious feature for the user's profile; in the case of users that are using the device in PORTRAIT mode, this feature may not be considered as important as others.

The set of chosen characteristics are very close to human behaviour specifics, like doing and correcting mistakes, using the service or the device as easy or efficient as possible. The study focuses on characteristics that distinguish the human user from the simulated one. Therefore, we took into consideration those characteristics at which software automation can't naturally exceed the human, like making mistakes or changing your mind.

4. Collecting behavioural data

In order to measure the characteristics related to the users' behaviour, we chose two different approaches, each one specific to the type of device that the user prefers:desktop or mobile.For the users that prefer to use their desktop to access the social platform, we use a desktop app that will embed the Facebook website inside a web-browser control.

The app is developed in the C# Windows Forms framework, this meaning that it will work only on Windows OS, nevertheless it can be easily ported to Linux or MacOS by using the open source WinForms Mono project component(Hiroshi & Timossi, 2006). An HTML and JavaScript approach could not be used because Facebook is blocking its site from being loaded inside *<iframe>* elements.

Each time the users starts the app, a full-screen form is opened containing a *webBrowser* control that loads the Facebook website. By using this additional layer, we can intercept user related events that we can process afterwards and obtain information about the user's behaviour characteristics that we are interested in.

For determining the scroll mode, we used two different events available in the .NET Framework, *MouseWheel* event on form level, to check if one is using the mouse to scroll, and *PreviewKeyDown*eventon the browser control level, to detect if one is pressing the up and down keys to navigate. The characteristic related to finding the typing mistakes is hardly measurable in this type of approach, first of all, because the WinForms component doesn't have any library related to typographical errors and, second of all, because we should have used a page scraping algorithm, that isknown to not be very reliable (Albrecht, Dean, & Hansen, 2005), to determine the HTML controls where the user can input data and listen to *textChanged* events on those.

The key used to delete text and the approach that one is using when it comes to correcting a typing mistake can be easily determined by using the same *PreviewKeyDown* event. The code used inside the handler for this event is available in Listing 1.All the events related to the studied behaviour characteristics are logged inside a file that can be processed afterwards (as it is or by uploading its content to a web service or distributed database) to compute the necessary metrics.

```
private void webBrowser1_PreviewKeyDown(object sender, PreviewKeyDownEventArgs e)
        {
if (e.KeyCode == Keys.Down || e.KeyCode == Keys.Up)
            {
                //the up and down keys are used to scroll
e.IsInputKey = true;
Console.WriteLine(Characteristics.ScrollWithUpDownKeys);
streamWriter.WriteLine(Characteristics.ScrollWithUpDownKeys);
return;
            }

            //...

if (e.KeyCode == Keys.Back)
            {
                //the user presses backspace to reach the error
nbOfBackspaceKeyPresses++;
Console.WriteLine(Characteristics.DeleteWithBackspaceKey);
streamWriter.WriteLine(Characteristics.DeleteWithBackspaceKey);
            }
        }
```

Listing 1. The code used for determining the type of scroll or the key used for deletion on the desktop app

For the users that prefer the mobile app provided by the social platform, we created our own Android mobile app that loads the mobile version of the Facebook site inside a *webView* widget. In this way, we can monitor the user's behaviour and determine the screen orientation in different situations like scrolling, reading a text or writing one.

For logging behaviour characteristics related data, we used the Realtime Database component from Firebase (the Google cloud platform). This component acts as aNoSQL distributed database (Mohajerani, et al., 2015) where all the data from the devices that we want to monitor can be stored and retrieved in real time. In Figure 4 a snapshot of the data stored in Firebase is shown.

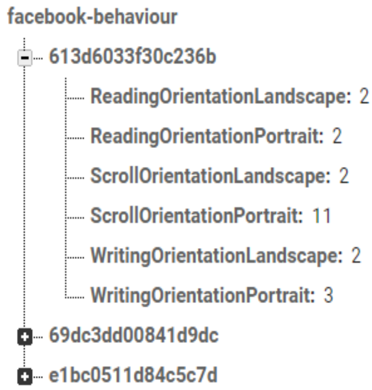


Figure 4. Part of the data stored inside the Firebase Realtime Database (the hash values represent phone unique ids)

In order to know when one is scrolling content from the Facebook Behaviour app that we have created, we placed the *webView* control inside a *ScrollView* and set an *OnScrollChangeListener*. If the orientation of the device is different from the one detected at the previous scroll we add a new entry in our distributed database. The code responsible for measuring this behaviour characteristic is visible in Listing 2.

```
scrollView.setOnScrollChangeListener(new View.OnScrollChangeListener() {
intoldOrientation= -1;
    @Override
public void onScrollChange(View v, intscrollX, intscrollY, intoldScrollX,
intoldScrollY) {
intorientation = getResources().getConfiguration().orientation;
if(orientation != oldOrientation) {
DatabaseReferencebaseRef = FirebaseDatabase.getInstance().getReference();
DatabaseReferencephoneRef = baseRef.child(androidId);
if (orientation == Configuration.ORIENTATION_PORTRAIT) {
DatabaseReferenceportraitScrollOrientation =
phoneRef.child(Behaviour.ScrollOrientationPortrait.toString());

portraitScrollOrientation.setValue(nbOfScrollsWithPortraitOrientation+1);
            } else if (orientation == Configuration.ORIENTATION_LANDSCAPE) {
                //...
        }
    }
});
```

Listing 2. Determining the orientation of the phone on scrolling for the mobile app

Because the Facebook mobile site that we load inside our browser control uses implicit Intents when the user clicks on a "read more" link or on an article, it was simple to determine the orientation of the screen used for reading news or articles. We just overrode the *onRestart* method from the main activity's lifecycle (which is called every time the user goes back to our app from a previously openedwebpage) and checked there for the orientation of the device.

Finally, to detect which type of orientation the user prefers when it comes to writing texts, we added an *OnGlobalLayoutListener* to the root view which, in our case, is the *ScrollView*. By checking the view's size and comparing it with the full size of the screen, we can determine if the keyboard is displayed or not. We chose this approach because there is no other event in the Android SDK that can be used to intercept the appearance of the keyboard.

5. Create the users profiles based on behavioural characteristics

Based on the survey done upon a number of 359 persons, a general profile can be concluded using clustering and similarity analysis. The first step in conducting an analysis upon the answers received from the responders is to transform the qualitative characteristics into quantitative ones containing an origin point.

The characteristics extracted can be grouped within the following main criteria:

- gender and age (G1);
- time spend on Facebook, Messenger and browser using mobile apps (G2);
- keyboard and ways of typing (G3);
- ways of responding to possible typing errors (G4);
- landscape and portrait mobile holding (G5).

The extracted groups of characteristics are initially analysed individually and after that there are created models of clusters, followed by a methodology for clustering the initial ones.

For the time spend on mobile apps referring to social media interaction, the questioner contains the following questions:

- Q1.1 How much time do you spend on Facebook mobile app?
- Q1.2 How much time do you spend on Facebook Messenger mobile app?
- Q1.3 How much time do you spend on Facebook using the browser?

All three questions contain a response within the following:

- <=1h
- between 1 and 2h
- between 2 and 3h
- between 3 and 4h
- >=4h

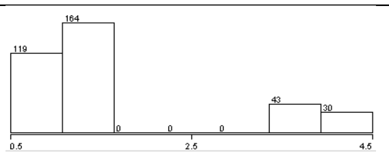
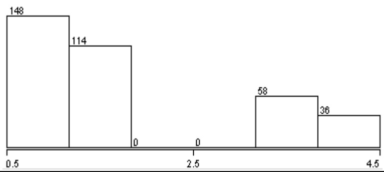
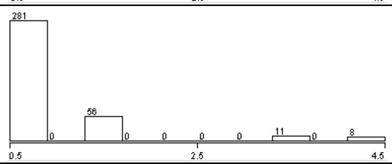
Table 1 contains a recodification of the following groups based on spending hours. The codification is done using the mean of the initial response interval.

Table 1. Recoding of responses given by the responders

Initial response	Response codification
<=1h	0.5
between 1 and 2h	1.5
between 2 and 3h	2.5
between 3 and 4h	3.5
>=4h	4.5

For the missing values, the standard for replacing was done using the average of the responses given by the others. Using Weka analysis software, table 2contains the descriptive statistics for the three questions.

Table 2. Descriptive statistics for the responses obtained for the first three questions

Q	Min value	Max value	Mean	Standard deviation	Chart
Q1.1	0.5	4.5	1.66	1.247	
Q1.2	0.5	4.5	1.71	1.392	
Q1.3	0.5	4.5	0.84	0.825	

In order to identify the relation between the three questions, table 3 contains the correlation values between those, 0 corresponding to no correlation between the variables and 1 being the total correlation.

Table 3. Correlation matrix between the responses for questions Q1.1, Q1.2 and Q1.3

	Q1.1	*Q1.2*	*Q1.3*
Q1.1	1		
Q1.2	0.580812	1	
Q1.3	0.193309	0.149972	1

The greatest correlation is between Q1.1 and Q1.2, meaning that the more time it is spent on Facebook mobile app, the more time is also spent on Messenger mobile app. While, in the opposite direction, the correlation between question Q1.2 and Q1.3 is very low, indicating that persons using Messenger and Facebook mobile app are not using the desktop browser version of Facebook. Starting from this initial analysis, clustering is conducted in order to identify groups of homogeneous users of Facebook on mobile devices. For that, using Weka software for K-means clustering algorithm under Euclidean distance evaluation, the following centroids were extracted. A splitting of 66% to 33% was used for the training and testing sets, resulting in Figure 5.

Cluster centroids:

		Cluster#	
Attribute	Full Data	0	1
	(356)	(107)	(249)
Q1.1	1.6601	2.8738	1.1386
Q1.2	1.7135	3.5841	0.9096
Q1.3	0.8399	0.986	0.7771

Figure 5. Centroids values for the three clusters formed using the input data extracted from the questioner

Cluster 0, containing a set of 107 responders is formed out of those who are using more the Facebook mobile apps available, generating a centroid of (2.87; 3.58; 0.98), while cluster 1, formed out of 249 responders, is characterized by a centroid of (1.13; 0.90; 0.77). The greatest differences are present for Q1.1 and Q1.3 between the two centroids, while for Q1.3 similar values are obtained. This indicates, as we also concluded from the initial descriptive statistics, that the time spent on Facebook using the browser has a lower standard deviation. As a first result, Q1.3 can be extracted from the general description of social media behaviour on mobile apps.

For the group keyboard and ways of typing, the following questions were asked with the survey:

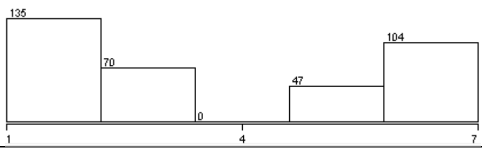
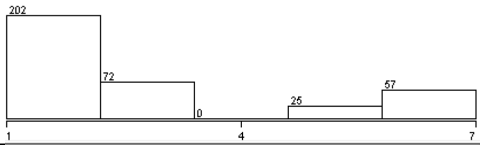
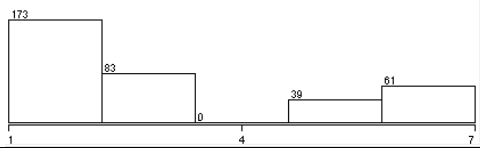
- Q2.1 How often do you use writing in apps and Facebook platforms?
- Q2.2 How often do you use the mouse in apps and Facebook platforms?
- Q2.3 How often do you use the keyboard in apps and Facebook platforms?

Each question has the following set of closed answers:

- <=2 times per day – coded with 1;
- between 2 and 4 times per day – coded with 3;
- between 4 and 6 times per day – coded with 5;
- more than 6 times per day – coded with 7.

After the initial descriptive statistics analysis, the following results from table 4 are obtained.

Table 4. Descriptive statistics for responses given to questions Q2.1, Q2.2 and Q2.3

Q	Min value	Max value	Mean	Standard deviation	Chart
Q2.1	1	7	3.67	2.50	
Q2.2	1	7	2.64	2.23	
Q2.3	1	7	2.93	2.26	

The correlations between the answers given for the three grouped questions are present in table 5. The greatest correlation is between the answers given for questions Q2.2 and Q2.3.

Table 5.Correlation matrix between the responses for questions Q2.1, Q2.2 and Q2.3

	Q2.1	*Q2.2*	*Q2.3*
Q2.1	1		
Q2.2	0.293963	1	
Q2.3	0.482231	0.664635	1

For the clustering algorithm, again the kMeans is used under the Euclidian distance evaluation, obtaining the following results, Figure 6.

```
Cluster centroids:
                              Cluster#
Attribute     Full Data           0          1
                  (234)        (93)      (141)
==============================================
Q2.1             3.6838      6.1398     2.0638
Q2.2             2.6667      4.2258     1.6383
Q2.3              2.906      4.7419      1.695
```

Figure 6. Centroids values for the three clusters formed using the input data extracted from the questioner

Cluster 0, containing a set of 93 responders is formed out of those who are using more in the Facebook app the mouse, keyboard and writing, generating a centroid of (6.13; 4.22; 4.74), while cluster 1, formed out of 141 responders, is characterized by a centroid of (2.06; 1.63; 1.69).
For the group of questions under landscape and portrait holding of the mobile, while using different apps, the following questions are present.

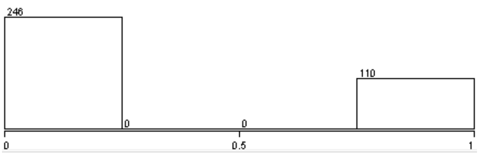
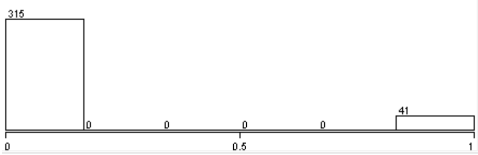
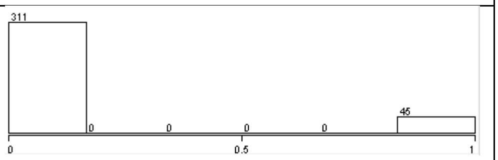
- Q3.1 How do you hold your mobile while trying to read a long text?
- Q3.2 How do you hold your mobile while trying scrolling on Facebook?
- Q3.3How do you hold your mobile while trying to write a text?

Do all three questions have the following available answers?

- Portrait – coded with 0;
- Landscape – coded with 1.

The coding with 0 for the Portrait version of holding the mobile device was done because of the consideration that portrait holding is the normal, standard way, while landscape mode is conducted intentionally whenever a user feels the need. For these questions, the following descriptive statistics presented in Table 6 were obtained.

Table 6. Descriptive statistics for responses given to questions Q3.1, Q3.2 and Q3.3

Q	Min value	Max value	Mean	Standard deviation	Chart
Q3.1	0	1	0.31	0.46	
Q3.2	0	1	0.11	0.32	
Q3.3	0	1	0.12	0.33	

From initial analysis, the way of holding the mobile device, regardless of situations described in the three questions, is more likely to be done in the normal, portrait holding position. The correlation between the answers of the three questions is present in Table 7.

Table 7.Correlation matrix between the responses for questions Q3.1, Q3.2 and Q3.3

	Q3.1	*Q3.2*	*Q3.3*
Q3.1	1		
Q3.2	0.349089	1	
Q3.3	0.404207	0.736602	1

The greatest correlation is present between Q3.3 and Q3.2, a value of 0.73, meaning that the holding position while using a scrolling of Facebook and typing is similar. For the clustering analysis, again the kMeans algorithm under the Euclidian distance is conducted, concluding with the values of the centroids found in Figure 7.

```
Cluster centroids:
                          Cluster#
Attribute     Full Data          0          1
                  (234)      (164)       (70)
=============================================
Q3.1             0.2991          0          1
Q3.2             0.1111     0.0366     0.2857
Q3.3             0.1282     0.0427     0.3286
```

Figure 7. Centroids values for the three clusters formed using the input data extracted from the questioner

Cluster 0, containing a set of 164 responders is formed out of those who are using significantly more the portrait holding position of the mobile device, generating a centroid of (0.00;0.03;0.04), while cluster 1, formed out of 70 responders, is characterized by a centroid of (1; 0.28; 0.32).

The next phase of the current analysis is combining the results obtained for the three sets of questions (Q1, Q2 and Q3). Each respondent was grouped within a cluster depending on the lower Euclidian distance between the centroid of each available cluster and the values of their responses. Figure 8 contained the re-clustering of those values.

Cluster centroids:

Attribute	Full Data (234)	Cluster# 0 (121)	1 (113)
ClusterQ1	0.312	0.6033	0
ClusterQ2	0.3974	0.7686	0
ClusterQ3	0.7009	0.7521	0.646

Figure 8. Re-clustering results based on the previous three clustering algorithms applied on the obtained data

Cluster 0, containing 121 responders, is formed out of those who were closer clustered within cluster 1 for question group 1, 2 and 3, while cluster 1 is formed out of 113 responders who were prior clustered within cluster 0, or close to it.
A different approach is consisted in using the raw initial data, not clustered data, obtaining the following results while clustering under K-means with Euclidian distance.

Table 8. Clustering results using the raw initial data

Cluster centroids:

Attribute	Full Data (234)	Cluster# 0 (82)	1 (152)
Q1.1	1.7308	2.5244	1.3026
Q1.2	1.7778	2.7561	1.25
Q1.3	0.8248	1.2317	0.6053
Q2.1	3.6838	6.1951	2.3289
Q2.2	2.6667	4.3171	1.7763
Q2.3	2.906	5	1.7763
Q3.1	0.2991	0.2561	0.3224
Q3.2	0.1111	0.122	0.1053
Q3.3	0.1282	0.1585	0.1118

We obtained two clusters using the DBSCAN (Density-Based Clustering Algorithm) algorithm available within Weka software. Figure 9 contains the descriptive statistics of mean and standard deviation of the two obtained clusters.

```
Cluster: 0 Prior probability: 0.3517

Attribute: Q1.1
Normal Distribution. Mean = 2.5244 StdDev = 1.4978
Attribute: Q1.2
Normal Distribution. Mean = 2.7561 StdDev = 1.5603
Attribute: Q1.3
Normal Distribution. Mean = 1.2317 StdDev = 1.1692
Attribute: Q2.1
Normal Distribution. Mean = 6.1951 StdDev = 1.392
Attribute: Q2.2
Normal Distribution. Mean = 4.3171 StdDev = 2.5609
Attribute: Q2.3
Normal Distribution. Mean = 5 StdDev = 2.1413
Attribute: Q3.1
Normal Distribution. Mean = 0.2561 StdDev = 0.4365
Attribute: Q3.2
Normal Distribution. Mean = 0.122 StdDev = 0.3272
Attribute: Q3.3
Normal Distribution. Mean = 0.1585 StdDev = 0.3652
```

```
Cluster: 1 Prior probability: 0.6483

Attribute: Q1.1
Normal Distribution. Mean = 1.3026 StdDev = 0.8663
Attribute: Q1.2
Normal Distribution. Mean = 1.25 StdDev = 0.9884
Attribute: Q1.3
Normal Distribution. Mean = 0.6053 StdDev = 0.4161
Attribute: Q2.1
Normal Distribution. Mean = 2.3289 StdDev = 1.7312
Attribute: Q2.2
Normal Distribution. Mean = 1.7763 StdDev = 1.5398
Attribute: Q2.3
Normal Distribution. Mean = 1.7763 StdDev = 1.3387
Attribute: Q3.1
Normal Distribution. Mean = 0.3224 StdDev = 0.4674
Attribute: Q3.2
Normal Distribution. Mean = 0.1053 StdDev = 0.3069
Attribute: Q3.3
Normal Distribution. Mean = 0.1118 StdDev = 0.3152
```

Figure 9. Descriptive statistics for the two obtained clusters

A prior probability was used for each cluster, a value of 35.17% for cluster 0 and 64.83% for cluster 1.

Based on our tests, the collected data was not enough in order to create a fail-proof user profile that can be used to recognize identity theft. We decided to “decorate” the user profile with some other indicators that are known to uniquely identify each individual like typing biometrics (Araújo, et al., 2005)(Feng, Zhao, Carbunar, & Weidong, 2013). Because not every time one connects to a social network they write something, we added also the time between different scrolls and the scroll length as characteristics(Clarke, Furnell, & Reynolds, 2002)(Peacock, Ke, & Wilkerson, 2004)(Weidong, Yang, Jiang, Yang, & Xiong, 2011).

6. Conclusions

User profiling in social media is a must concern in the area of mobile security along with the web. Having an increase in the sensitive data stored within mobile devices and also the social media platforms available for mobile devices, a user profiling is needed for proper validation of the person who is using the mobile device. The objective of this research paper is to determine a set of characteristics used within a classification model in order to differentiate the behaviour of users while scrolling, typing and reading. For this, the survey conducted with 20 questions referring to different methods of using a mobile app for social media extracted the base of the study, having as results qualitative and quantitative data of a prior selection of the set of characteristics.

The statistical results of the survey were then integrated with several collecting behaviour methods implemented for mobile and web versions of social media apps usage.

Forming five sets of characteristics mainly focused on gender and sex, time spent of on Facebook, Messenger and browsers, keyboard and ways of typing, ways of responding to typing errors and ways of mobile holding in landscape and portrait format, generated multiple analysis in terms of factor impact upon classification functions.

Using correlation matrix clustering algorithms, the study has shown that two main categories of users can be created. One category is formed out of the users that have a background knowledge when it comes to mobile usage and tend to optimize their work utilizing either landscape or portrait placement of the mobile device, correction of typing in less time and modifications, directly related to the time spend on different social media mobile apps. The other group is formed out of users that interact harder with the characteristics offered by a mobile device, results also expressed in the obtained data.

Even though the results are not statistically significant in order to obtain an accurate model for classification, it offers a base point for future work when it comes to user profiling in social media apps. Clustering algorithms with k-Means and Euclidian distance generated the best combination while creating the two groups of users, 35% of the responders being placed in one group, while the rest of 65% being placed in the group of experienced one.

Future work will be focused on creating a wider set of characterises that are mapped to certain users, not just groups of users, and, along with a training and testing methodology, a classification function will differentiate the real user to the others that are not having the same behaviour.

Acknowledgment

This paper presents results obtained within the PN-III-P1-1.2-PCCDI-2017-0272 ATLAS project ("Hub inovativ pentru tehnologii avansate de securitate cibernetică / Innovative Hub for Advanced Cyber Security Technologies "), financed by UEFISCDI through the PN III –"Dezvoltarea sistemului national de cercetare-dezvoltare", PN-III-P1-1.2-PCCDI-2017-1 program.

REFERENCES

[1] **Albrecht, C., Dean, D. L. & Hansen, J. V. (2005),** ***Marketplace and Technology Standards for B2B e-Commerce: Progress, Challenges, and the State of the Art***. *Information & Management, 42*(6), 865-875;

[2] **Araújo, L. C., Sucupira, L. H., Lizarraga, M. G., Ling, L. L., T. & Yabu-Uti, J. B. (2005),** ***User Authentication through Typing Biometrics Features***. *IEEE transactions on signal processing, 53*(2), 851-855;

[3] **Boyd, D. M. & Ellison, N. B. (2007),** ***Social Network Sites: Definition, History, and Scholarship***. *Journal of computer mediated Communication, 13*(1), 210-230;

[4] **Clarke, N. L., Furnell, S. M. & Reynolds, P. L. (2002),** ***Biometric Authentication for Mobile Devices***. *Proceeding of the 3rd Australian information warfare and security conference;*

[5] **Enge, E. (2018, December 5),** ***Mobile vs Desktop Usage in 2018: Mobile Takes the Lead***. Retrieved from Stone Temple: https://www.stonetemple.com/mobile-vs-desktop-usage-study/;

[6] **Feng, T., Zhao, X., Carbunar, B. & Weidong, S. (2013),** ***Continuous Mobile Authentication Using Virtual Key Typing Biometrics. Trust, Security and Privacy in Computing and Communications (TrustCom)****, 2013 12th IEEE international conference on;*

[7] **Heaven, D. (2018, October 1),** ***Massive Facebook Data Breach Left 50 Million Accounts Exposed***. Retrieved October 25, 2019, from NewScientist: https://www.newscientist.com/article/2181099-massive-facebook-data-breach-left-50-million-accounts-exposed/;

[8] **Hiroshi, N. & Timossi, C. (2006),** ***Mono for Cross-platform Control System Environment***. Lawrence Berkeley National Laboratory;

[9] **Hunt, A. & Gentzkow, M. (2017),** ***Social Media and Fake News in the 2016 Election***. *Journal of Economic Perspectives, 31*(2), 211-36;

[10] **Kayacik, H. G., Just, M., Baillie, L., Aspinall, D. & Micallef, N. (2014),** ***Data Driven Authentication: On the Effectiveness of User Behaviour Modelling with Mobile Device Sensors***. *Proceedings of the Third Workshop on Mobile Security Technologies (MoST).* San Jose: IEEE Computer Society Security and Privacy Workshops;

[11] **Mohajerani, S., Syed Ali Hashim, M., Rihawi, R.-A., Ahmed, B., Akhlaque, N. B. & Kamal, R. Y. (2015),** ***A cloud-based system for real-time, remote physiological monitoring of infants***. *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT).*

[12] **Peacock, A., Ke, X., & Wilkerson, M. (2004).** ***Typing patterns: A key to user identification***. *IEEE Security & Privacy, 2*(5), 40-47.

[13] **Sasse, M. A., Brostoff, S., & Weirich, D. (2001).** ***Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security***. *BT Technology Journal, 19*(3), 122–131. doi:https://doi.org/10.1023/A:1011902718709

[14] **Statista (2018, November),** ***Market Share Held by the Leading Social Networks in the United Kingdom (UK) as of November 2018***. Retrieved November 2, 2019, from Statista. The Statistics Portal: https://www.statista.com/statistics/280295/market-share-held-by-the-leading-social-networks-in-the-united-kingdom-uk/;

[15] **Webb, I. G., Pazzani, M. J. & Billsus, D. (2001),** ***Machine Learning for User Modeling. User Modeling and User-Adapted Interaction****, 11*(1-2), 19–29. doi:https://doi.org/10.1023/A:1011117102175;

[16] **Weidong, S., Yang, J., Jiang, Y., Yang, F., & Xiong, Y. (2011),** ***Senguard: Passive User Identification on Smartphones Using Multiple Sensors. Wireless and Mobile Computing, Networking and Communications (WiMob)****, 2011 IEEE 7th International Conference on;*

[17] **Zuckerberg, M. (2012, October 4),** ***Newsroom*****.** (Facebook) Retrieved August 29, 2019, from https://newsroom.fb.com/news/2012/10/one-billion-people-on-facebook/.