Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

Determination of Potential Criminals in Social Network

Ramazan CESUR[1,*], Eyüp Burak CEYHAN[2], Ayten KERMEN[1], Şeref SAĞIROĞLU[1]

[1] *Computer Engineering Department, Gazi University Engineering Faculty, 06570, Ankara, Turkey*

[2] *Computer Engineering Department, Bartın University Engineering Faculty, 74110, Bartın, Turkey*

Article Info

Received: 17/05/2016
Revised: 27/11/2016
Accepted: 16/01/2017

Keywords

Twitter analysis
Mmachine learning
Cyber-crimes
Social network
Crime analysis

Abstract

This paper proposes a new approach to determine potential criminals by performing content analysis on Tweets. The analysis is done with the help of machine learning technologies and big data analysis. In this study, we have utilized from the MLP algorithm. A dataset consisting of 384 words are used to make the classification process. Dataset consist of two classes that are organized crime and cyber-crimes. In the analysis process, date, time, location values of sent twitter sharings are also used. Criminals can be detected with a success rate of 71,61%. Also, the developed system identifies potential criminals who commit an offence of organized crime and cybercrime with the world's most widely used social media. User can use the system to get accurate results for scanning potential criminals with analyzing their sharings or scanning keywords to reach potential criminals. In addition to this property, user can get results that are more accurate with narrowing the content screening with location and date information. It is thought that the proposed system might help to find criminals and the security forces can easily detect them by the developed software.

1. INTRODUCTION

Internet has become indispensable with the advanced technology. Many people are able to do their business with the internet connection. While these developments affect positively the welfare of human life, they have also brought some negative consequences. Especially the information shared in almost every field of life via social networks that is getting more common constitutes a huge source for abuse. Organized criminal gangs find both supporters and followers by using the information on social network profiles, and also they can easily determine their victims from there. The criminals, who use these networks that the information is collected from everywhere, commit new crimes with the information they obtained from these environments [1]. The mass that are affected from virtual environment is large and the domain is very wide. It poses a threat to the one of the basic principles of the society such as life assurance, guarantee of the goods, the security of religion and the assurance of mind and generation occurred by imposing opinion. Social media becomes a part of daily life of society. With users from almost every environment, social media ensures the connection between the worldwide users and it is spreading rapidly. These features increase the use of social media as a potential crime tool and new crime areas become accessible for individuals and groups who are subject to commit crimes [2].

Studies [2, 3] shows that the increase in the number of crimes committed in cyber environment continues and crime committing ways become much more special, also the effect level they caused is quite high. This situation has encouraged researchers to find solutions in this field. In order to prevent and fight against these crimes and these negative effects increased with the technological developments simultaneously, there are developments in the field of security.

According to data from the Statistics Institute of Turkey, computer and internet usage rate of the 16-74 age groups are 53% in Turkey in 2014. 16-24 age groups are the group with the highest rate of Internet use. Internet use in all age groups is higher in men than in women. The first three months of 2014, the users used internet to: 78% use social networking sites, 74% read the news, read a magazine or newspaper, 67% get information about goods and services, 58% download and play games, movies music, 53% follow their

[*] Corresponding author, e-mail: ramazancesur3@gmail.com

e-mails. In 2013-2014 period, the usage rate of internet for communication increased 11% considering 2012-2013 period. In the 12 months period between April 2013 and March 2014, while 51% of users who orders goods on the internet, bought clothing and sportswear, 27% of them bought household and 24% bought technological tools [4]. These statistics indicate that the internet is an indispensable part of daily life. Having a utilization rate of social media with 78%, shows that social media is very effective and it is appropriate to use these networks for security purposes.

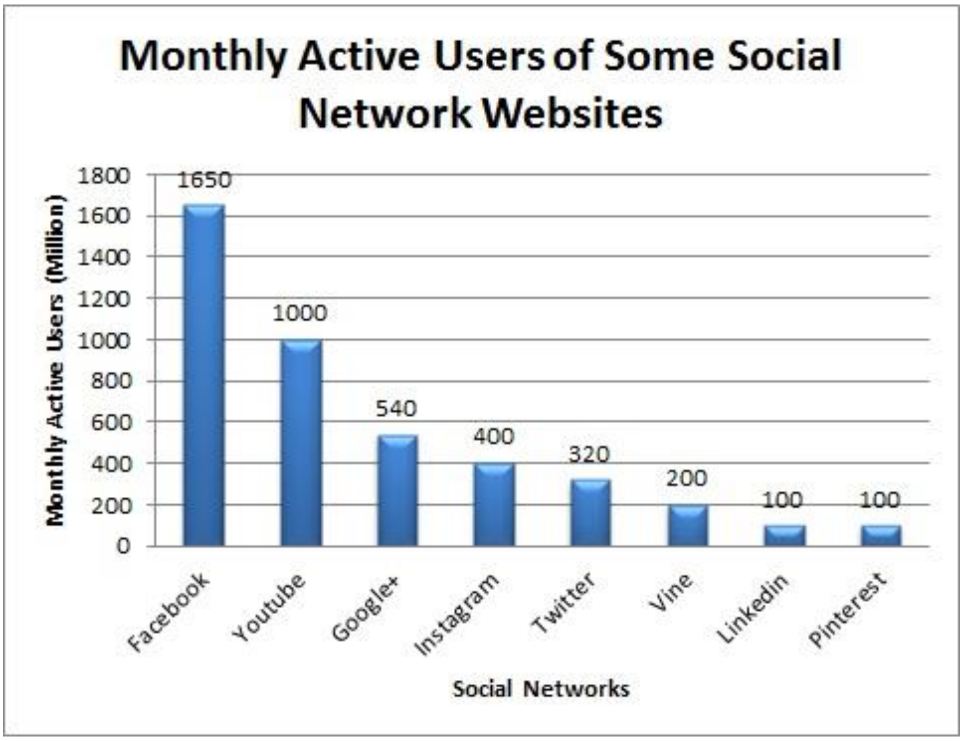


Figure 1. *Monthly active users of some social network websites [8] (values 1 represent by 1 million account)*

Figure 1, which consists of May 2016 statistics, shows the number of members of the social media sites [8]. Facebook accounts can be seen in first place with approximately 1.7 billion members. Due to Facebook's privacy and security policy, it is more difficult to examine user sharings, so using Twitter data is more appropriate because everybody can see the sharings. Examples such as the Arab Spring and Gezi Park events in Turkey show clearly the impact of Twitter on masses. It is also easier and more efficient to analyze Twitter sharings consisting of 140 characters.

This paper consists of five parts. Literature review is in second part. In the third part of the article, the information about the basic working algorithm, working steps, software flowchart and the quality of the data used in the developed system was given. In the fourth part, user interface of the system and the analysis of the results by determining the suspect were described in detail and the information about data sets in database was given. In fifth part, it was mentioned about the success obtained from the system used and the results obtained from the algorithms tried to use in the system were evaluated and compared. In the last part, it was mentioned that how to increase the success of the system and the improvements could be done in future studies by deliberating the necessity and the benefits of the system proposed.

2. LITERATURE

Day by day, it becomes inevitable that people get a habit to share their data and thoughts often or their social network usage to spread scientific messages and harmful thoughts [5]. This shared data, not only are a problem for society, but also they constitute a source, which is ready to use and limitless, for criminals who go after the personal data. This common use shows that the people don't need a specific place to commit a crime and their presence in this large network is enough to commit a crime. The guilty must be present at the scene in any kind of crime. Insulting, emptying bank accounts, stealing internet accounts belonging to people (social networking sites, email and other interactive accounts), stealing personal data (images, videos and other documents belonging to individuals), even and even stealing the most secret information of the government can easily be done by means of a device which has a connection with the internet from miles away or a smart phone [6].

M. S. Gerber, made a study predicting crime with kernel density estimation [18]. This study examines issues of linguistic and statistical modelling spatiotemporally Twitter sharings by identifying issues for discussion in the United States. In this study, an increased crime prediction performance than the standard approach is obtained.

Researchers and several commercial organizations have performed analysis from Twitter data on several subjects such as health, natural disasters, epidemics, TV programs, sports, advertising and politics [7, 9-15].

In 2013, Edward Crook analyzed random Tweets by classifying them into the fields such as TV, sports, music, health, literature and politics in order to determine the expectations and detect the thoughts of British female and male Twitter users about the brands. As a result of the research, there was not a notable difference between male and female users. "Brand types Mentioned on Twitter" shows the distribution percentage as shown in Figure 2 [7].

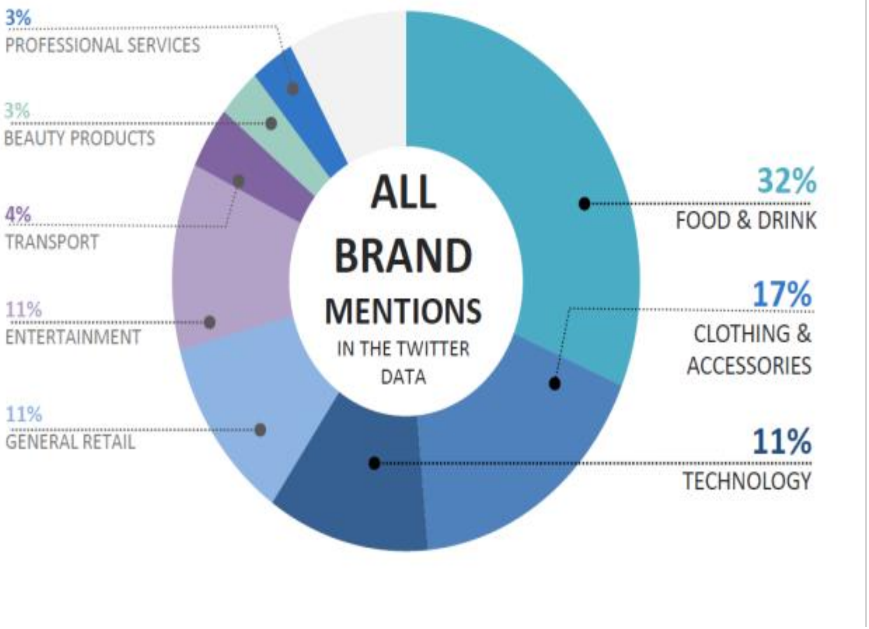


Figure 2. *Brand types Mentioned on Twitter [7].*

In another analysis study, Gupta and his friends conducted a study to determine the reliability of popular events on Twitter. They developed a system that analyses simultaneously the large data sets in a few minutes by classifying Twitter shares with various functions according to their belonging in real time. As a result of study performed by two real data sets used, the event graph-based optimization has better results than the basic reliability approach and traditional classifier approach [9].

In 2011 in Japan, Acar and Muraki performed a study based on detecting a first aid call in disasters like tsunami, by the aim of ensuring the crisis communication about S.O.S call and people's condition and the condition of the area of natural disaster [10]. Also about the same subject in 2010, there is a study of Sakaki and his friends [11]. They investigate the real-time interaction of events such as earthquakes in Twitter and propose an algorithm to monitor tweets and to detect a target event.

In another study about natural disasters, Twitter usage was investigated during and after the Haiyan typhoon in Philippines. It was researched how the users used Twitter and for which purpose during the natural disasters. It was noticed that Twitter was the most commonly used source as an information source with a rate of 43,4%. To memorialize the events it was also determined that it was used with a rate of 32% [12].

In a study performed in 2014, Twitter shares of Turkish and Dutch users were analyzed. This research, which was performed to detect the diversity of point of views and point of views of minority, provided the opportunity to see the different aspects of the diversity of point of views. It was also provided to see the location of the minority opinions in Turkish and Dutch cultures [13].

In a study in which the contents of tweets tweeted by Presidential candidates in 2012 were analyzed, the dialogues of candidates with their followers and their political transparency in this dialogues were examined. It was noticed that the tweets about the economy (unemployment, tweets related to budget and tax) were one of the best three subjects on tweets for all candidates. Although they are important for the voters, health, foreign affairs and social subjects are the least shared titles of the tweets. As a result of the study, it was seen that the candidates didn't use Twitter to have a conversation with the voters and to answer their questions. This study revealed that the candidates were not sufficiently transparent on Twitter pages [14].

In the 2014 World Cup, during US team and the opponent's matches, large data analysis was performed with the tweets of sports fans in real time. In the study, the emotions of the fans were analyzed by examining the tweets that the fans sent during the matches. When the US team scored, positive emotions, and when the opponents scored, negative emotions showed rising. In the matches between the opponents, distinctive sense transitions were not seen [15].

Chan, Cho and Yang from the University of Virginia was conducted a study [19] to determine the time and place of a specific crime in the United States. They found that the model they created was more successful than the core density estimate by testing their ability to predict future crimes. This study reveals that Twitter data has an important potential for crime detection.

3. PROPOSED APPROACH AND SYSTEM

Proposed system is created for Security Forces in order to detect the people who share tweets containing actual and criminal threats in Twitter. The system offers information about the membership names of Twitter accounts, usernames, definitions that the users made for themselves, Tweet contents they sent and the dates and places that the shares were made. These information are about people who share criminal tweets.

The developed system can be used to find the suspects who have a Twitter account or to detect the potential criminal by performing a content screening. In addition, when determining a potential criminal, the suspect's identity, gender, address, or other social media web addresses linked, if any, the date of the membership information are also accessible.

The suspicious Twitter user's data chosen from result list which is brought according to the entries taken from the user, is transferred to the second interface. Shares of suspect taken by system are separated to perform semantic analysis. Conjunctions, prepositions, punctuations and internet connection are discarded. If more than 50% of data quality of the user that is shown in Table 1 is reached, share contents are subjected to training and testing processes. If 50% of the data quality cannot be reached, data processing is made by erasing the unreached qualities and by passing into the second step of machine learning. As a result of the process, a matrix is created by the data. Weights are given to the words and sentences according to their meanings. Then, by performing 7 times and 10-times cross-validation process to the 384 words of obtained data, the data is classified into two groups. The result is returned by detecting the crime type of suspect such as organized crime or cyber-crime according to the result. 0 values seen in Result column show cyber-crimes, 1 values show organized crimes.

Table 1. *The data used in the application.*

ID	Location	Tweet	Day	Month	Time	Result
1	Ankara	Robbery	08	03	15	1
2	İstanbul	Metasploid	23	08	02	0
3	Ankara	Hack	18	11	21	0
4	Adana	Heroin	15	07	23	1
… **…**	… …	… …	… …	… …	… …	… …
383	İzmir	Spoofing	30	04	17	0
384	Ankara	Cocaine	04	06	11	1

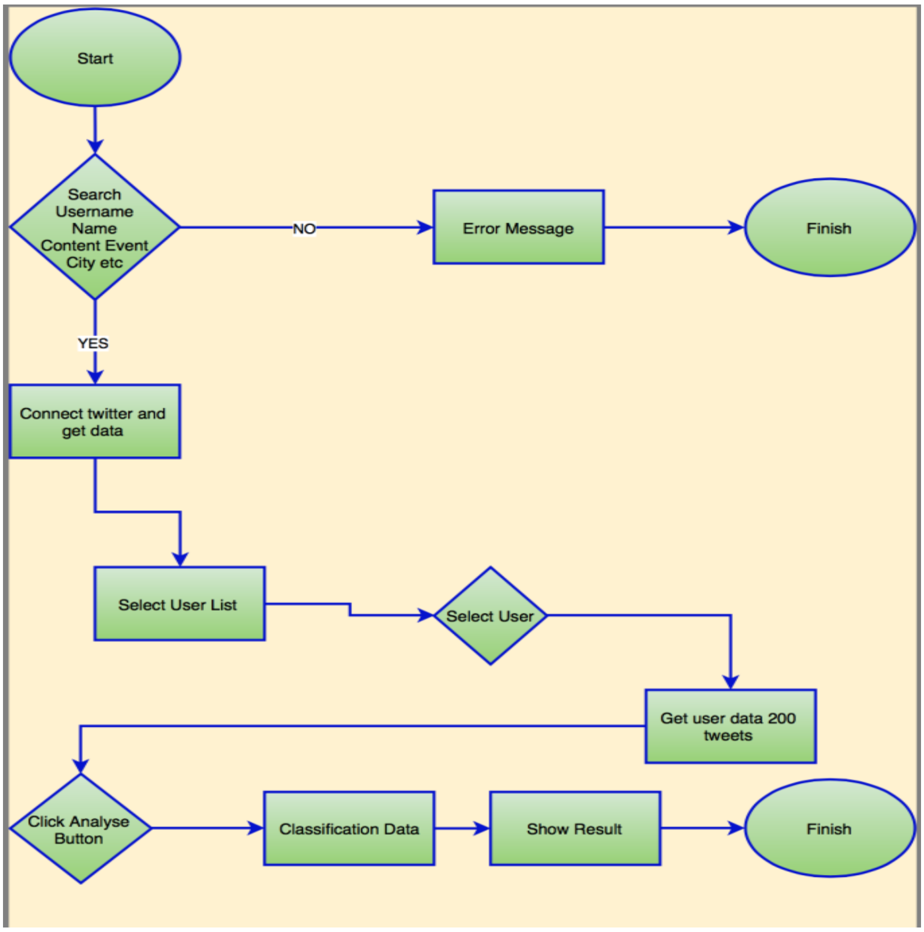


Figure 3. *Software Flowchart*

The software flowchart belonging to the proposed system is shown in Figure 3.

The logic of the system can be summarized in four steps:

1. According to the information received from the user, person or content scanning is performed.
2. Suspect is detected by listing the most popular results.
3. Tweets and user data of the detected suspect are collected.
4. It is found that the person is a potential criminal or not, and if so, the crime type that he/she involved, by scanning and analyzing the shared contents.

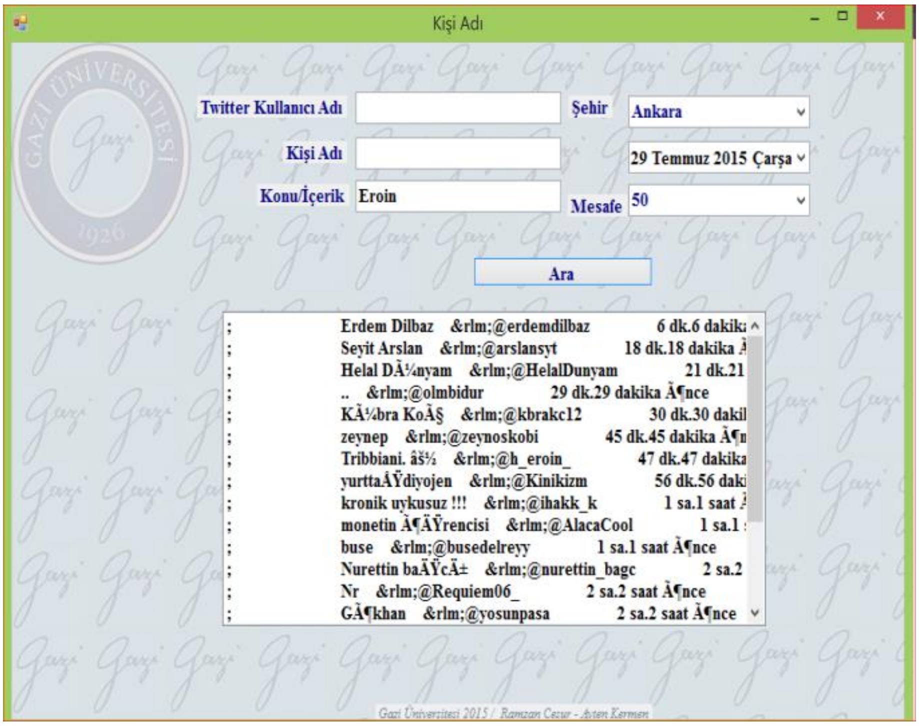
The developed system identifies potential criminals who commit an offence of organized crime and cybercrime with the world's most widely used social media. User can use the system to get accurate results for scanning potential criminals with analyzing their sharings or scanning keywords to reach potential criminals. In addition to this property, user can get results that are more accurate with narrowing the content screening with location and date information.

4. DEVELOPED SYSTEM AND RESULTS

The developed system consists of two main modules. First, Twitter account information of suspect, which is found by searching a specific subject, is taken in usable data format. Secondly, the tweets of determined person are classified into two titles by using specific algorithms and it is determined that the shared contents are involved into crime, and if so, which crime that is.

4.1 Suspect detection and data extraction

The first module of the system is to detect the suspect and to extract the data about suspect to classify. In this section, system does the searching process by taking words or phrases from suspect. Figure 4 shows the user interface belonging to this part of the system.



Twitter Kullanıcı Adı:

Kişi Adı:

Konu/İçerik: Eroin

Şehir: Ankara

29 Temmuz 2015 Çarşa

Mesafe: 50

```
;      Erdem Dilbaz   &rlm;@erdemdilbaz        6 dk.6 dakik:
;      Seyit Arslan   &rlm;@arslansyt        18 dk.18 dakika Ã
;      Helal DÃ¼nyam   &rlm;@HelalDunyam          21 dk.21
;      ..   &rlm;@olmbidur         29 dk.29 dakika Ã¶nce
;      KÃ¼bra KoÃ§   &rlm;@kbrakc12           30 dk.30 dakil
;      zeynep   &rlm;@zeynoskobi         45 dk.45 dakika Ã¶n
;      Tribbiani. âš½   &rlm;@h_eroin_         47 dk.47 dakika
;      yurttaÅŸdiyojen   &rlm;@Kinikizm         56 dk.56 daki
;      kronik uykusuz !!!   &rlm;@ihakk_k         1 sa.1 saat Ã
;      monetin Ã¶ÄŸrencisi   &rlm;@AlacaCool          1 sa.1 :
;      buse   &rlm;@busedelreyy          1 sa.1 saat Ã¶nce
;      Nurettin baÄŸcÄ±   &rlm;@nurettin_bagc          2 sa.2
;      Nr   &rlm;@Requiem06_        2 sa.2 saat Ã¶nce
;      GÃ¶khan   &rlm;@yosunpasa          2 sa.2 saat Ã¶nce
```

Figure 4. *Suspect Search Interface*

As seen in the interface of the system in Figure 4, searching process is performed based on the username, person's name or the content of the share. The searching process can be narrowed down by selecting city, date and distance information. By selecting the distance, it is determined that Twitter contents or users will be searched how much km^2 away from the city center entered in the interface. This makes it possible to obtain more results that are specific. Searching process covers all users around the world, primarily the ones in Turkey. 100 results are listed in the results of the search.

If Twitter user name of the suspect is known, the search is done by typing a user name into "Twitter User Name" field. Because of Twitter usernames cannot be identical, Twitter information and shares of the suspect can be directly reached. Then the obtained information is analyzed in the background of the system.

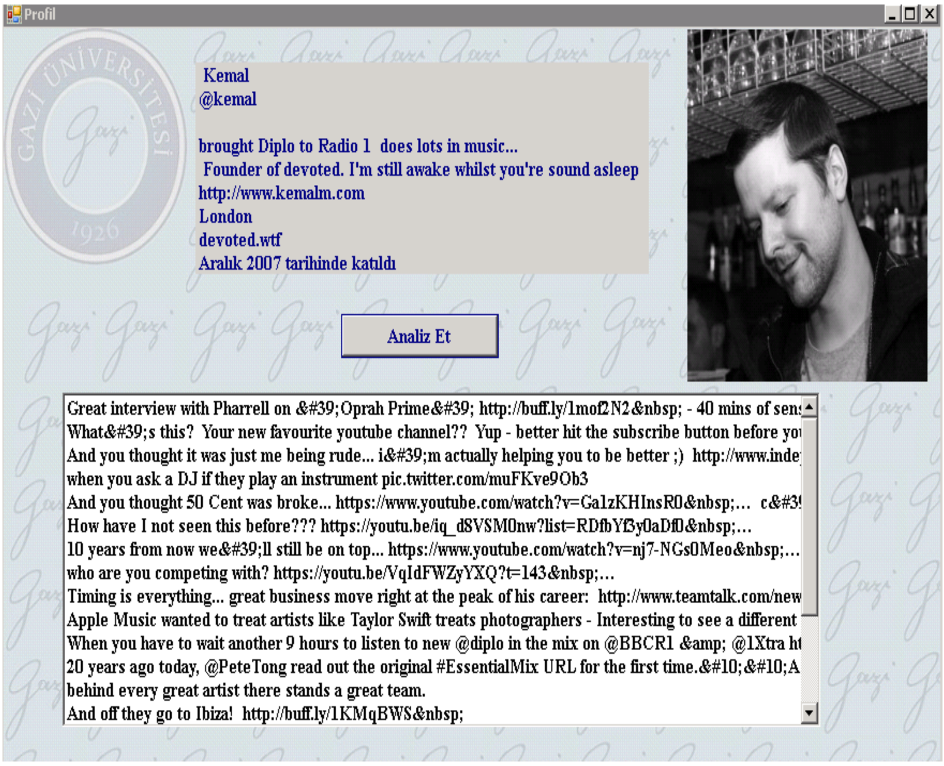
If it is not certain that suspect has a Twitter account, search process is performed by entering the person's name into "Person name" field. Many users can have the same name. With the searching process, most popular results that have the same name with suspect are shown. After determining the number of results, Twitter account of suspect is detected amongst the listed people and the information is analyzed by transferring the data to the second stage.

In cases there are no suspects for a type of crime but there is a criminal suspicion about the subject, content search can be done by taking data from "Subject/Content" field. Determination of the city and the date is very useful when searching content. The owners of the most popular Twitter shares are listed when searching content about the subject.

In this part of application, Twitter account of the suspect is detected. After taking one of the data such as username, person's name or content, data, which is suitable for the entries, is listed on the screen by pushing "Search" button. In the user's list, which is suitable for entries, person's name and Twitter username are shown. When any of the listed result is chosen, the user is directed to the analysis interface shown in Figure 4.

4.2 Classification and analysis process

After completing the first stage of application and detecting Twitter account of the suspect, the information about suspect is transferred to the analysis screen shown in Figure 5.



Kemal
@kemal

brought Diplo to Radio 1 does lots in music...
Founder of devoted. I'm still awake whilst you're sound asleep
http://www.kemalm.com
London
devoted.wtf
Aralık 2007 tarihinde katıldı

```
Great interview with Pharrell on &#39;Oprah Prime&#39; http://buff.ly/1mof2N2&nbsp; - 40 mins of sens
What&#39;s this?  Your new favourite youtube channel??  Yup - better hit the subscribe button before yo
And you thought it was just me being rude... i&#39;m actually helping you to be better ;)  http://www.inde
when you ask a DJ if they play an instrument pic.twitter.com/muFKve9Ob3
And you thought 50 Cent was broke... https://www.youtube.com/watch?v=Ga1zKHInsR0&nbsp;...  c&#3
How have I not seen this before??? https://youtu.be/iq_d8VSM0nw?list=RDfbYf3y0aDf0&nbsp;...
10 years from now we&#39;ll still be on top... https://www.youtube.com/watch?v=nj7-NGs0Meo&nbsp;...
who are you competing with? https://youtu.be/VqIdFWZyYXQ?t=143&nbsp;...
Timing is everything... great business move right at the peak of his career:  http://www.teamtalk.com/new
Apple Music wanted to treat artists like Taylor Swift treats photographers - Interesting to see a different
When you have to wait another 9 hours to listen to new @diplo in the mix on @BBCR1 &amp; @1Xtra ht
20 years ago today, @PeteTong read out the original #EssentialMix URL for the first time.&#10;&#10;A
behind every great artist there stands a great team.
And off they go to Ibiza!  http://buff.ly/1KMqBWS&nbsp;
```

Figure 5. *The interface on which the analysis process is performed.*

As seen in Figure 5, name, surname, contact information, city, date of membership or if so, the connections belonging to the other internet sites or accounts and person's profile picture, which belong to Twitter account determined in the first stage, are extracted. The most crucial point of the application is the extraction of Twitter shares belonging to the person. Last 200 shares of the person, which are accessible to the public, are listed. When "Analyze" button is pressed, the contents of the listed shares are classified and analyzed.

One of the important factors that increase the success of the software in classification process is the quality of the extracted data after preliminary processing. System is processing the information of location, where the share is made by latitude and longitude, date in the format of day/month/year and time of share by extracting them except from the information shown to the user in the background. By this way, it is intended to achieve optimal results by machine learning from the data.

The classification process is performed according to the database of predetermined words related offenses. There are two types of data in the database. First of these is the words about the cyber-crime. There are 195 words in the database on cyber-crime.

The second data set contains 189 words related to organized crime. There are a lot of offenses within the scope of organized crimes. In the advanced version of the application, these crimes will be more detailed and classified.

The developed system uses the information that public Twitter users share. The information that the user does not wish to share, is not used and personal rights are not violated. When the application is given for the usage of law enforcement, the information that can be accessed by law enforcement agencies within legal limits, may also be used.

4.3. Classification results

The system classifies the results for organized crimes and cyber-crimes by analyzing the Twitter sharings of potential criminals. Organized crimes consist of narcotic, grab, fraud and many similar crimes.

Two dataset that contain crime words were used in classification process. Attributes of the data used was shown in Table 1, before.

KNN, MLP and SVM algorithms were used for classification When creating the system, various classification algorithms were used with looking similar studies and their successes. KNN (K Nearest Neighborhood) algorithm was first tried out from the success of a study analyzing popular events on Twitter [9]. It has been understood that classical classification algorithms will not be sufficient to solve this problem because the success achieved using the KNN classification algorithm is not sufficient. In [16], MLP (Multi Layer Perceptron) algorithm achieves high success with a system that labels music by semantically web-based information management. It was then integrated into an artificial neural network system that was built on the MATLAB platform using the MLP algorithm. In [17], Support Vector Machine (SVM) algorithm was used because it provides 91% success to estimate the political orientations of Twitter users. The highest accuracy obtained in our system was 71,61% with using 10-fold cross validation technique and MLP algorithm.

In our system, the highest classification success ratio was obtained by using the MLP algorithm with a rate of 71,61%. The classification accuracies obtained from our study were shown in Figure 6. Also, mean absolute error, F-measure and ROC area results are shown in Table 2.

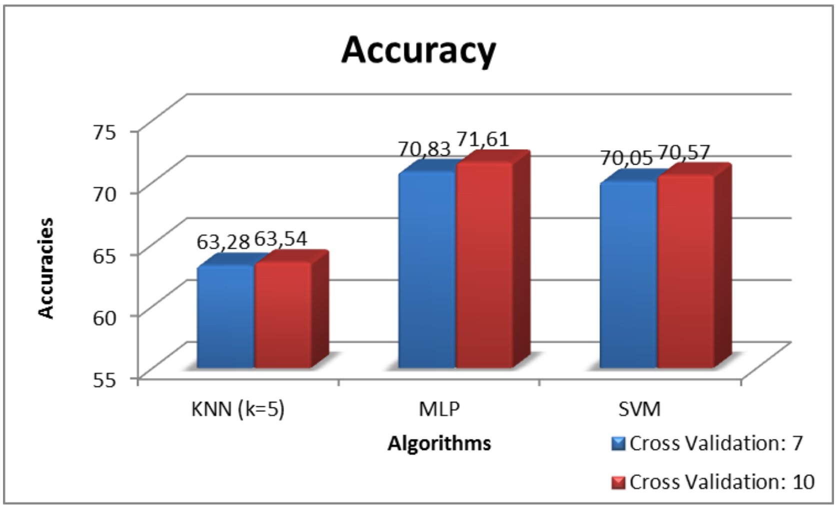


***Figure 6**. Accuracies with using different algorithms in proposed system.*

Table 2. *Mean absolute error, F-measure and ROC area results.*

Algorithm	Mean Absolute Error	F-Measure	ROC Area
KNN (CV=7)	0.43	0.63	0.66
MLP (CV=7)	0.34	0.71	0.78
SVM (CV=7)	0.30	0.70	0.70
KNN (CV=10)	0.43	0.64	0.67
MLP (CV=10)	0.34	0.72	0.78
SVM (CV=10)	0.29	0.71	0.71

In Table 3, aim, method and used dataset of literature studies similar to topic of this paper were shown.

Table 3. *Aim, method and used dataset of literature studies*

Paper	Aim	Method	Dataset
[9]	Determination of the reliability of the popular events	KNN	D2010 dataset: 288 events D2011 dataset: 250 events
[16]	Information Management on Semantic Web	MLP	>40.000 songs (10 different types)
[17]	Determining political trends of Twitter users	SVM	200 most popular domains
This study	Estimation of potential criminals	MLP	384 tweets

5. CONCLUSION

With the technological developments in our age, people who commit crimes on the internet, can easily communicate to each other by using social networks. Social networks have an indispensable position in the lives of each age group and each fraction. This has made people share their personal data, opinions and information more often or made inevitable the people's usage of social networks as a tool to reach a social aim. This shared data constitute a very large source for criminals, who go after the personal data especially.

The age of social network users is falling steadily. Especially children and young users are quite open to the influence of organized crime gangs. Expressing the information and opinions means accessible quality to determine a target group for illegal politic groups or terrorist organizations.

Therefore, there is a need to develop new tools and methods to combat these crimes. In order to fight effectively against cyber-crime, data in digital media, it is the most powerful material to prove the crime. In the fight against cyber-crime, the evidence obtained through forensic means earlier, it will accelerate the finalization of the case and possible damage will be minimized. The application developed with this aim, uses Twitter data in order to detect the criminal and crime gangs. It analyses Twitter shares of the suspect and classifies them. It can also detect the previously unidentified criminals by analyzing the content shared on Twitter. To identify the person who commits a crime on social networks, to contribute to the elucidation of real-life forensic cases and to detect the social network users that damages to the community in several aspects, are intended.

In order to use the digital data obtained via forensic tools as evidence by juridical authorities, the forensic tools that the evidences are obtained from, must have the features such as capable of identification, prediction, being repeatable, being verifiable [6]. Because of the system developed in the scope of this study provide these features; they can be used as forensic tools.

In the process of criminal investigation, share the experience obtained with these kinds of studies and the developed tools with national security units can block the crime in different places and also deterrence can be achieved.

Being unofficial of the profiles created on social networks, it also allows the malicious person to hide his identity. This case is suitable for the crime gangs to reach large population without leaving much evidence, to direct people and to use their information for improper purposes.

This study is one of the big steps to follow closely changes in crime event, to make detections more accurate and faster, to be more effective in interfering and taking precautions and to secure the internet environment as actual life. It will make easier to detect the mass and global movements such as Arabic Spring, especially which started and spread in social media, and to take the necessary precautions. By fixing provocateur people and their gangs, it will be able to prevent the events which could adversely affect the public welfare.

Currently, classification is made for two different types of crime in the system. In the future studies, it is planned to achieve classifications that are more successful by adding new crime areas to the classification and by expanding the crime databases.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors

REFERENCES

[1] Cinar, B., “Sosyal Medyanın Örgütlü Suç İşlemede Rolü”, JOBEPS: International Journal of Business, Economics and Political Science, 1(2): 79-102, (2012).

[2] Acar, A. and Deguchi, A., “Culture and Social Media Usage: Analysis of Japanese Twitter Users”, International Journal of Electronic Commerce Studies, 4(1): 21-32, (2013).

[3] Signorini, A., Segre, A. M. and Polgreen, P. M., “The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic” PLoS ONE, 6(5): 1-10, (2011).

[4] Internet: “TUIK Statistics”, http://www.tuik.gov.tr/PreHaberBultenleri.do?id=16198, Access Date: 17.05.2016.

[5] Letierce, J., Passant, A., Breslin, J. and Decker, S., “Understanding How Twitter is Used to Spread Scientific Messages”, Web Science Conference (WebSci10), Raleigh, NC, USA, 1-8, (2010).

[6] Özdemir, A., “Adli Bilişim Araçları”, 1st International Symposium on Digital Forensics and Security, Elazig, Turkey, 1-4, (2013).

[7] Crook, E., "The Twitter Landscape", A Brand watch social insights report, (2012).

[8] Internet: "Social Media Active Users by Network [INFOGRAPH]", http://www.thesocialmediahat.com/active-users, Access Date: 17.05.2016.

[9] Gupta, M., Zhao, P. and Han, J., "Evaluating Event Credibility on Twitter", 12. SIAM International Conference on Data Mining, California, USA, 153-164, (2012).

[10] Acar, A. and Muraki, Y., "Twitter for Crisis Communication: Lessons Learned from Japan's Tsunami Disaster", International Journal of Web Based Communities, 7(3): 392-402, (2011).

[11] Sakaki, T., Okazaki, M. and Matsuo, Y., "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors" in Proceedings of the 19th International Conference on World Wide Web, New York, USA, 851-860, (2010).

[12] Takahashi, B., Tandoc, E. C. and Carmichael, C., "Communicating on Twitter during a Disaster: An Analysis of Tweets during Typhoon Haiyan in the Philippines", Computers in Human Behavior, 50: 392-398, (2015).

[13] Bozdag, E., Gao, Q., Houben, G. and Warnier, M., "Does Offline Political Segregation Affect the Filter Bubble? An Empirical Analysis of Information Diversity for Dutch and Turkish Twitter Users", Computers in Human Behavior, 41: 405-415, (2014).

[14] Adams, A. and McCorkindale, T., "Dialogue and Transparency: A Content Analysis of How the 2012 Presidential Candidates Used Twitter", Public Relations Review, 39(4): 357-359, (2013).

[15] Yu, Y. and Wang, X., "World Cup 2014 in the Twitter World: A Big Data Analysis of Sentiments in U.S. Sports Fans' Tweets", Computers in Human Behavior, 48: 392-400, (2015).

[16] Kolozali, S., Barthet, M. and Sandler, M., "Knowledge Management on the Semantic Web: A Comparison of Neuro-Fuzzy and Multi-Layer Perceptron Methods for Automatic Music Tagging", 9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012), London, England, 220-231, (2012).

[17] Conover, M., Goncalves, B., Ratkiewicz, J., Flammini, A. and Menczer, F., "Predicting the Political Alignment of Twitter Users", IEEE Third Inernational Conference on Social Computing (SocialCom), Massachusetts, USA, 192-199, (2011).

[18] Gerber, M.S., "Predicting crime using Twitter and kernel density estimation", Decision Support Systems, 61: 115-125, (2014).

[19] Chen, X., Cho, Y., Jang, S.Y., "Crime Prediction Using Twitter Sentiment and Weather", Systems and Information Engineering Design Symposium, Charlottesville, USA, 63-68, (2015).