IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

Social Networks and Education Level Relationship

Rudi Miran Babayi[1], Adem Ozyavas[2], Ali Gunes[3]

Istanbul Aydin University, İstanbul, Turkey[1-3]

Abstract: This work deals with estimating the education level of social media users. There may be different reasons for inferring new information from social media data. Security and economical ones are the top ones. Data mining – especially Idea Mining – on social networks have become more widespread in the international literature. Even though the number of studies done domestically have been increasing, it is not enough. This work aims to take a small step towards finding out the education level of Twitter users. The last part of this work shares the experience we had out of this research for future studies.

Keywords: Social Networks, Data Mining, Twitter.

I. INTRODUCTION

The need for communication seems to be ever-present. . With the widespread use of the Internet, it has shifted to web-based applications. As a result, the data now are being saved in digital form. This led to the increasing need for saving large amounts of data in new database systems [1].

Web-based applications are not designed for communication purposes only. Social media users have used them to have good time as well.They allow their users near and far to communicate with ease with the touch of keys[2].

People aorund the world take advantage of soical media platforms such as the most popular Facebook, Twittter, and Youtube allowing them to share. This, in turn, leads large amounts of data to be saved in database systems. Social media database systems generally allows access from outside. There has been a trend in recent years to use these databases for analyzing and profiling the social media user. New applications are being developed to turn this data into useful information. One of the methods used for extracting information is data mining. Data mining makes it possible to profile social media users and determine the link between user and their social groups. It, therefore, helps to predict and guess what users feel about a certain topic based on the shares they have made. In this work, we evaluated Twitter users' shares to help us make inferences about them using data mining. As we mentioned before, our purpose with this study is to infer the education level of Twitter users based on the shares they have made. We confined our language to Turkish. We determined the users' education level based the on the frequency of the key words they use in their shares. Becuse the Internet's ease of accessibility to masses web-based applications are used more and more for communication purposes. This led a shift in the communication preferences of users from old traditional ways to the new way. What made these web-based applications more applealing is the fact that they have become a tool for spending time with what interests them. Thus, they evolved into a new way for people to have fun together. Using these social media applications, people express themselves about what and how they feel about political issues, form groups to share with others[3].

Social Network Analysis (SNA): Producing meaningful information based on the analysis of relationships obtained from social media using some methods. It is also defined as networks of relationships and their effects on the social structure [4,5].

II. MODEL

We used the Twitter database of its users' shares to infer the education level of Twitter users. Estimating the education level of Twitter users is important in proper identification of categories of people which helps in terms of what products should be offered to them. A brief look at the literature reveals that in classifying the groups of Twiter users can be done based on the key words they used in their shares. For instance, social media users are categorized as happy or unhappy based on the key words they used in their shares with the key words picked imply the fact the users are happy ones. The analysis of such data results in the binary outcome so that either the users belong to a category or not. The binary outcome is a natural result of data mining based on the key word and text analysis. We would need more data if we wanted to determine the levels of happiness. Therefore, the overwhelming approach in the listerature are the binary results of whether a property exists or does not exist.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

In this study, we first defined what edcuation level we are interested in and attempted to estimate if Twitter users are at/above that education level or below that one. We determined the higher eduction as the level of education. Therefore, the results of this study should help us identify the Twitter users with higher eduction degree based on the shares they have had during a certain period of time. The above infernece is made based on the key words they used in their shares. If they use certains key words implying that they had higher eduction degree (1), they are recorded as so. If they do not use those key words, then they are considered not to have a higher educcation degree (0).

Analysis of tweets are performed based on the fact that whether preselected key words occur in those tweets or not. Twitter users are categorized whether they have a higher eduction degree or not. We prefer to show the results using vector model [6]. We used a binary matrix where each element of the matrix show the presence or absence of the poperty we are interested in analyzing. This method is heavily used in similarity analysis.

Each tweet denotes a vector which is presented below.

V={above/at higher eduction, below higher eduction} showing presence or absence.
Any tweet associated with (1,0) shows one with a higher eduction degree whereas (0,1) denotes the absence of such a degree.

For instance,

I cannot feel Sunday because I have to study #ales

The above tweet uses the key word ales (a test for graduate schools). The Twitter user sharing the above message will qualify to be with a higher degree V=(1,0) so that she will be included in the group with a higher degree. New inferences can also be made after encoding data in the binary vector. For example, we can obtain the frequency of Twitter usage of such target users with a higher degree by obtaining the ratio of tweets with key words by the target users to all tweets by all users during a certain period. Therefore, we can obtain such a ratio for the Twitter users with a higher degree. Below is the mathematical expression showing the ratio:

$$\frac{\sum_{i=0}^{n-1} tweet_{kw}(i)}{\sum_{i=0}^{n-1} tweet_{any}(i)}$$

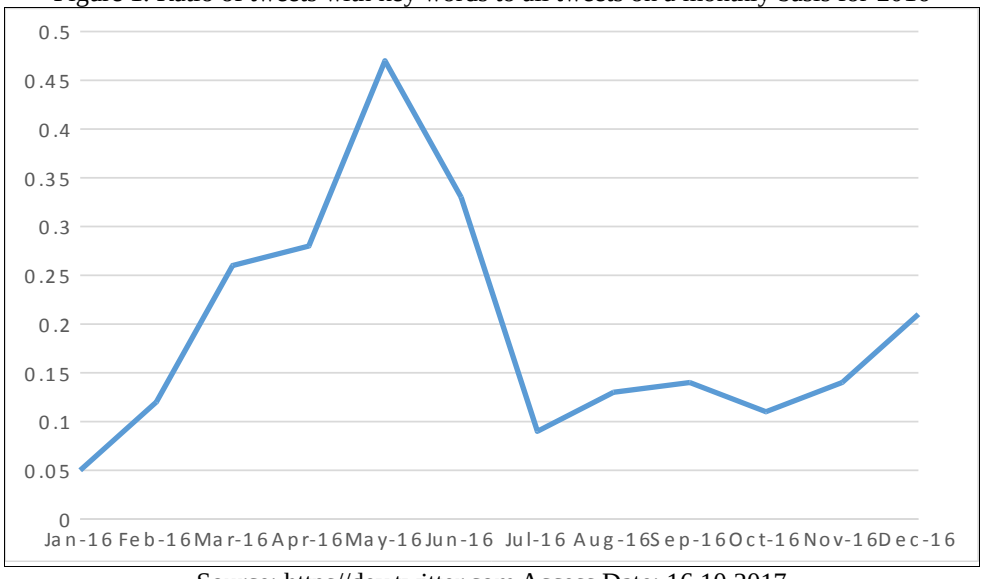
where

$tweet_{kw}(i)$ means 1 if the i^{th} tweet uses the keyword, 0 otherwise,
$tweet_{any}(i)$ means 1 whether the ith tweet uses the key word or not,
n is the number of all tweets inspected

III. APPLYING THE MODEL

In Figure 1, we present the the ratio of tweets with the key words to all tweets for the year 2016 from Turkey.

Figure 1. Ratio of tweets with key words to all tweets on a monthly basis for 2016



Source: https//dev.twitter.com Access Date: 16.10.2017

DOI 10.17148/IJARCCE.2017.61217

IJARCCE

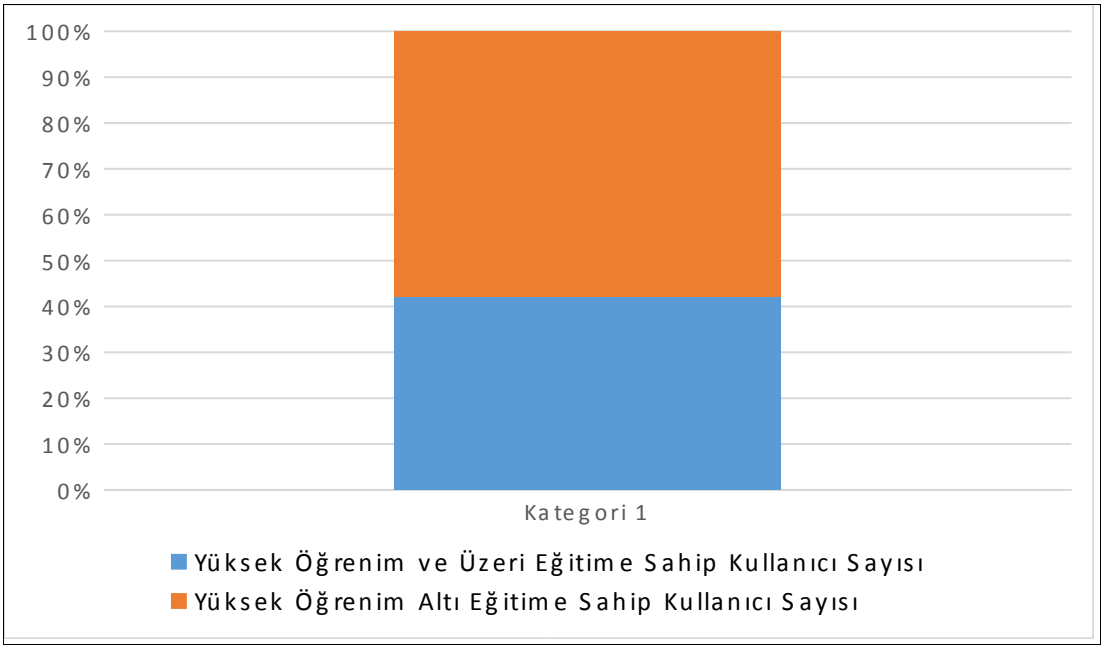
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 12, December 2017

A quick look at Figure 1 will show the ratio of tweets with key words to all tweets on a monthly basis for 2016. Figure 1 shows a low ratio of tweets from users with higher degrees to all tweets. The highest ratios are in May and June which is almost half of all the tweets. The average ratio of users with higer education degrees is around 10% of all the tweets.

We observe that the ratios are shaped based on the exam dates for the gradudate schools. As a result, the dates for ALES and KPSS (exam to recruit employees to goverment instutions) exams are the times during which this ratio raises. Similarly, TUS (medical graduate school specialization exam) and DUS (dental graduate school exam) also had an effect in the ratios. The ratios tend to go higher in favor of the users with higher degree again at the end of the year because of exams to have a place at the Ministry of Justice, Supreme Court of Public Accounts, etc... We conclude that the ratios go higher around the exam dates prepared for people with higher education degrees.By analyzing the contents of the Tiwtter messages, it is possible to find out the ratio of users with higher education degree to all Twitter users. Figure 2 shows this ratio.

Figure 2. The ratio of Twitter users with higher education degrees to all Twitter users (2016)



According to the data we analyzed 42% of all Twitter users have higher education degrees.We observe that the analysis in Figure 2 is higher compared to the content analysis we did in Figure 1. The results in Figure 2 do not agree with the statistics released by the internationally trusted organizations. Wearesocial is one of those organizations. They released a report in 2016 using the method of data mining based on key word search. Our results are not in accordance with this report.

In Digital in 2016 Global Overview report, there are 15 million Twitter users in Turkey. 63% of those users have higher education degrees which is higher than the percentage of that other social media users.Therefore, the results obtained from this study show percentages below the actual percentages. This may be because of the method of analsysis, the way we handle data, etc... Since our method of analysis of data mining based on (key) word search is simple but with some disadvantages to it. Ours is one of those cases where the search for a key word search using data mining falls short of a comprehensive analysis.Even though this simple approach may have some disagreement with the data released by internationally trusted companies, it still can give a general trend. Final general discussions on our analysis and shortcomings of our approach are provided in Section IV.

IV. CONCLUSION AND DISCUSSION

We utilized the idea mining method to estimate the education level of Twitter users based on some key words. Similar to other works in the literature, we attempted to answer whether a certain property exists or not of a phenomenon. We looked into the data to determine as closely as we can whether Twitter users have higher education degrees or not. We conclude that our results show we can rely on the method we used to estimate the education level of Twitter users. There is a discrepancy between our findings and reports from the internationally trusted companies. There coulde be several factors contributing to this disagreement. The hardness of our method results from the fact that our analysis

drastically depends on picking the proper key words. Therefore, the shortcomings related to picking the proper key words, the period of analysis, etc.. all affect our results and thus the discrepancy between our findings and those of internationally trusted companies'. Still, this work can be considered a small step to further the domestic work. We believe that our method can be used together with the other methods (classification between groups and people) to help increase the confidence in the results of future research.

REFERENCES

[1]. Berendt, B. , Hotho, A., Stumme, G.(2010), "Bridging the gap - Data mining and Social network analysis for Integrating Semantic Web and Web 2.0", *Journal of Web Semantics*, 8(2-3): 95-96.
[2] Huisman, M., Van Duijn, M.A.J., (2011), "The SAGE Handbook of Social Network Analysis", 578-600.
[3]. W. Koehler,(2001), Information science as "Little Science": The implications of a bibliometric analysis of the Journal of the American Society for Information Science. Scientometrics, 51(1):117-132.
[4] Kuduğ, H.(2011),(2011), "Sosyal ağ analizi ölçütlerinin iş alanlarına uyarlanması", Yüksek lisans tezi, Ege Üniversitesi Fen Bilimleri Enstitüsü.
[5] S. Wasserman, J. Galaskiewicz, (1994), Advances in Social Network Analysis: Research in the Social and Behavioral Sciences, Sage Focus Editions, Thousand Oaks, California, 91320.
[6] Lu, L., Zhou, T., (2011), "Link Prediction in Complex Networks: A survey", Physica A 390, 1150-1170.