



DATA MINING: Analysis of a Social Data Visualization Web Site (Instagram)

Manar Saad¹, Aljawhara Mohammad², Rahaf Hamri³ and Ghada Fahad⁴

College Of Business Administration, Imam Abdurrahman Alfaisal. Saudi Arabia, Dammam

¹*210005790@iau.edu.sa*

ABSTRACT

in the recent years, everyone tends to use social media applications that affect our life and interests. One of the most effective and influence application is Instagram. It is a photo and video sharing that allow people communicate together around the world. Also, it shows people lifestyle in interesting way that why Instagram become most popular social network platform. Instagram draws many of people by showing many different categories such as traveling, food, fashion, and other. This paper seek to analyze some kind of data in Instagram and make it easily when people want to search on it by using two important techniques, which are Rabid Miner and Tableau. Moreover, these data will be filtered based on the most famous account that have highest followers, which are celebrities and make some operator on it. Finally, to conclude these data in better and meaningful way, using visuals and graphs.

Keywords: *Social, Instagram, Celebrities, Tableau, Rabid Miner, Operator.*

1 INTRODUCTION

Nowadays social media become one of the most things that affect people life in many ways. Social media analysis makes the communication easier and faster between people around the world also help people gather sufficient information about what they look for. Social media analysis helps in make huge improvement in the methods used to gather the useful data [2]. Social media applications are many such as Instagram, Twitter and Snapchat. Instagram is a camera company that allows people to share pictures and videos that show people lifestyle in interesting way. Each day millions of people enjoying and capturing their moments with their families and friends through this application. Instagram is the first in social media that growing quickly currently and the first in user involvement. Moreover, Instagram users are more than 400 million. In addition, the pictures that uploaded are more than 100 million for each day and likes or comments more than 10,000 in seconds that shows users are highly active in this application. Instagram attracts many people by showing different categories and interests such as traveling, shopping, food and celebrities account that creates many followers inspired by them. One of the Instagram features is Insta story that users share their pictures and videos and last for only 24 hours then delete automatically. Other feature is users can record videos life [6].

2 VISUAL ANALYSIS OF SOCIAL MEDIA DATA

In the social media domain, users can make content in many forms-including video, images, text, and geospatial data- that is freely available. These data can be used for many purposes such as, by Companies to improve their business, by public health to monitor outbreaks of infectious diseases, or by celebrities to promote their businesses or albums if they are singers. Using social media data poses many challenges, these data are large in volume and are transmitted in high frequency flows also, data are multimodal, often ambiguous in its content, and interactive visualization stimulates analysts' creativity and discovers their background knowledge. The application of visual analytics highlights many benefit social media data by this integrated approach. Social medial include all media formats in which group of user interact to produce and share information in a distributed and parallel process [4]. The most popular example includes instagram (photo and video sharing). The most common feature for instagram is that users can form interest groups or other types of communications (follower in Instagram).

In addition, social media produce vast amount of data that contain valuable information in different context. For example, analyst can use such data to track opinions about new products and services [4]. As well to track the most famous account that has highest number of followers and many more.

Visual analysis system for large and complex data integrates the discovery of computational knowledge with interactive visualization. The discovery of computational knowledge uses effective means of data mining to form pattern of large data sets. For example, grouping data into cluster or predict class label for invisible data. Visualization includes useful visual representation of data, such as graphs and diagrams [4].

3 RAPID MINER

Immediate Data Analytics tool that does not involve any programming code, but it provides many functions that applied on entered data to make cleansing, modeling, filtering and others [3].

a) As mentioned above, Instagram is one of the most popular social networking platforms that users use it to share their moments by posting a photo or a video. In fact, after doing some research it became clear that the most famous accounts that have highest followers are celebrities compared to sports and fashion accounts. Many celebrities have an account on Instagram to share their important moments in their lives to their followers and fans and the content they place let their fans feel closer and more connected to the stars they like. Moreover, most of them promote some ads that is a low-cost way but generate high revenue. Our data has 100 of different categories, which are celebrities, sports, media, entertainment and fashion. After we have entered our data into Rapid Miner program and using Filter Examples operator, the program filtered category 1 and the result is 75 celebrities out of 100 as shown on Figure 1. This operator works after the user defined the condition and then the operator returns those examples that match the given condition.

After analyzing our data, we decide to focus on attributes that are useful and eliminate the rest which are ER and Iposts on Hashtag by using aggregate operator, which is like group by in SQL.

Then after we eliminate the unuseful attributes, we want to arrange our data by using sort operator to make either ascending or descending order for input data set.

Most international singers have an account on Instagram to share their new releases and their latest albums for their fans and they are considered as the most influential among celebrities in other fields. Our data support this fact by showing that most of the famous celebrities on Instagram are musicians after using Filter Example operator on category 2.

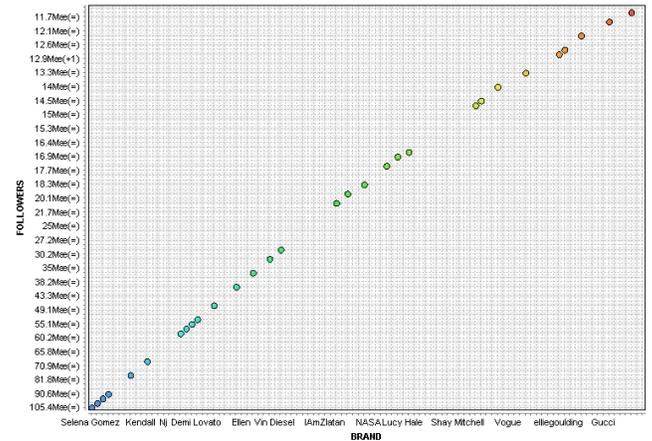


Fig. 1. First operator result

b) Then, we wanted to illustrate basic graph that representing a sample of each segment of celebrities and how many it uses Instagram. By using a select attributes operator and defining only categories1, categories2 and followers. After that, we used sample operator to stratify the data.

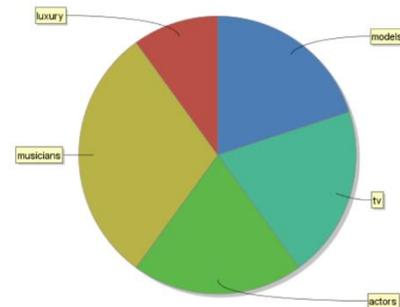


Fig. 2. Second operator result

c) Some cells in our data do not have any value, which may affect the accuracy of the result. Therefore, we use in Rapid Miner Program Filter Example operator to remove empty cells. After that, we want to group some segments together that are like each other so after analyzing our data, we pre-determined number of clusters that calculate the closeness between pairs of items which are 4 and the measure type is nominal measures. Figure 3 shows the result after using two operators.

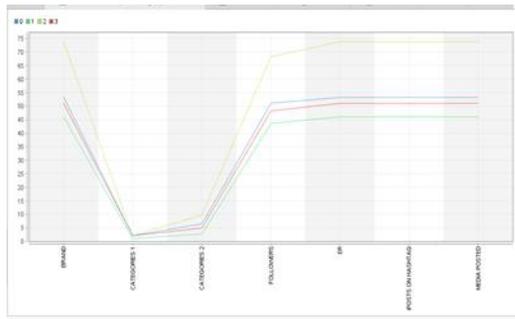


Fig. 3. third operator result

4 TABLEAU

A data mining application mainly focus on visuals and graphs that developers design it to conclude their data in meaningful way [5].

Nowadays Instagram consider as an important social media platform, which is a way for influencers to communicate, influence and transmit their mission to audiences and followers. For instant celebrities prefer to share their posts, videos and photos through Instagram. Moreover, celebrities receive high responds from audiences [7].

Figure 4 shows most popular categories in celebrities' accounts, splits into nine subcategories (actors, athletes, entrepreneurs, fashion, media, models, musicians, political and tv).

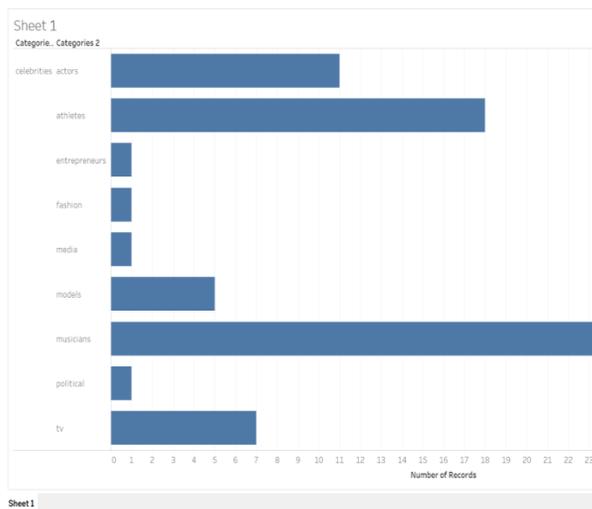


Fig. 4. celebrities categories

5 CONCLUSION

This project talked about social media analysis in Instagram area and how social media affect people around the world. Social media have been helping in increase communication and save people time by reducing the distances. Instagram is a camera company that provides the opportunity for people to share pictures and videos that show their unique lifestyle. The data gathered about Instagram was about the users, likes, comments, shares, following, followers and the active for each day. The Instagram statistics that have been collected and analyzed for different areas was update for 2018. Instagram application data are analyzed by using two programs RapidMiner and tableau. RapidMiner is software that for data preparation and used in different sectors such as education and training and tableau used for better recognizing and understanding the data analyzed.

6 REFERENCES

- [1] Conosquare Index Brands. (2018). Check out Top Brands doing best on Instagram #indexiconosquare. [Online] Available at: <https://index.iconosquare.com/> [Accessed 10 Apr. 2018].
- [2] Industrial & Management Engineering, IIT Kanpur. (n.d.). [ebook] p.4. Available at: https://www.iitk.ac.in/tkic/workshop/Big-Data/ppt/social_media_analytics.pdf [Accessed 14 Apr. 2018].
- [3] Rapid miner.com (2018). Operators. [Online] Rapid miner Documentation. Available at: <https://docs.rapidminer.com/latest/studio/operators/> [Accessed 12 Apr. 2018].
- [4] Schreck, T. and Keim, D. (2012). Visual analysis of social media data. Computer, [online] 46(5), pp.68-69. Available at: <http://ieeexplore.ieee.org/document/6389686/> [Accessed 15 Apr. 2018].
- [5] Tableau (2018). Tableau. [Online] Tableau. Available at: <https://www.tableau.com/> [Accessed 12 Apr. 2018].
- [6] worldtracker.academy (2018). Getting Started with Instagram Guide: Wordtracker. [Online] Wordtracker.com. Available at: <https://www.wordtracker.com/academy/social/visual/introduction-to-instagram> [Accessed 14 Apr. 2018].
- [7] Wright, K. (2015). Social Media and Celebrities: The Benefits of a Social Media Presence. [Online] Digitalcommons.calpoly.edu. Available at: <http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1093&context=jourisp> [Accessed 14 Apr. 2018].