



Social Media Analytics using Visualization: Snapchat

Noura Alsulaim¹, Noura Alkhalaf², Mashael Alghamdi³, Raghad Alolaqi⁴ and Hind Almughaidi⁵

^{1, 2, 3, 4, 5} Department of Management Information Systems, Imam Abdulrahman Bin Faisal University (IAU), Dammam, Kingdom of Saudi Arabia

ABSTRACT

Today, social media plays a crucial role in the life of the modern community. Many people use different social media means for different purposes on a daily basis, from communication with friends and family to promotion and of products. One of the biggest social media platforms is Snapchat where people share a big amount of data every day. With the increasing use of social media, the need to analyze the transmitted data appeared. Social media analysis aimed to identify tendencies and patterns of social media platforms that may be used for many purposes, from identification of the most popular industries to determination of the content that is most resonates to the audience. Social media data is a so-called raw data that includes all the activity of all the users of one or another social media platform. It should pass several processes before it can be analyzed and evaluated. Throughout these processes, data are freed from all the noisy data and is visualized to make it easier for analysis and evaluation. Tools such as Tableau and Rapid Miner were used to analyze and evaluate Snapchat's data for eight industries, and then data about the top ten accounts for each industry collected according to the number of views. The most viewed industry in Snapchat has been identified using Rapid Miner through K-Means algorithm. Other Rapid Miner's operators such as normalize, clustering and decision tree operators were used to evaluate data and make the prediction analysis for future tendencies.

Keywords: *Data Mining, Snapchat, Data Visualization, Tableau, Rapid Miner.*

1 INTRODUCTION

Now a day's millions of people use social media on a daily basis, which is a huge network that produces a large number of data. These data occupy approximately 30 percent of it is storage capacity. Redundancy and complexity are the major challenges for data analysts to deal with this high frequency data. As its volume continues to grow, it is obvious that the importance and necessity of data analytics and visualization cannot be underestimated [1].

Social Media Analytics is the process of collecting data produced from social networking platforms and extracting knowledge and information from these data using advanced tools of data mining. This information helps to make the right business decisions, learning customers' behavior, and it also useful to predict future events [1]. Social Media Analytics Process includes three steps: capture, understand, present. Presenting data visually helps to understand data by placing it in a visual context such as graphs, through which it can detect patterns, trends, and relationships that may not be detected in the textual form [2].

Snapchat is a social media app that visualize data for sharing multi-media, such as pictures and videos that viewed for a specific length of time before becoming inaccessible. The company reported 100 million daily active users worldwide positioning it as one of the fastest growing applications in the world [3]. Snapchat produces a large amount of data that may contain valuable information to be used by analyst for many purposes, for example tracking the most famous accounts and the industries in which the users are most interested in, also for marketers to engage with users, and determine the type of content that resonates with followers. In order to benefit from these big raw data, and by using tools such as Tableau and Rapid Miner, this paper analyzes and evaluates Snapchat's data. Through K-Means algorithm and decision tree operator in Rapid Miner program for eight different industries, comparing the number of views in each industry, and predictive data mining of the potential future number of Snapchat's users. Finally, it presents findings visually in graphs and diagrams, to show the results in a better and easier understanding way, using Tableau.

Thus, The Primary Research Objectives Are:

- A. Collect, monitor, analyze, summarize, and visualize Snapchat's data. By using advanced data mining tools (Rapid Miner and Tableau).
- B. Identify the industry on which the majority of Snapchat's users are interested in, to enable marketers easily choose predefined audiences to help them reach specific Snapchatters and reach people based on their online and real-world interests and behaviors.

2 LITERATURE REVIEW

This part discussed the two major concepts of this study, i.e. Social Media Data Analytics, and Data Visualization. Firstly, a brief discussion of the work of Social Media Data Analytics is provided, followed by a discussion on the methods and results of Data Visualization.

2.1 Social Media Data Analytics

Social media data is the raw data that results from individuals' activities on social media. These data track the users' participate in different contents and channels, so it is considered as a good source of raw information [4]. Social Media Analytics is the way to get benefit from these raw data, by extracting useful patterns and information, through three processes: *capture*, *understand*, *present*. In the *capture* process, the data gathered from various sources, preprocess the data and extract pertinent information from the data. *Understand* process contains remove noisy data, and perform advanced analytics. Then *present* the findings from the understand stage [5]. Social media analytics provides an answer to questions about people activities, such as:

- What types of content makes audiences click/share?
- What are the top-converting contents?

The answers of these questions are important to make the right business decisions, enable to identify which channel that can increase the revenue of the business and know the potential customers.

2.2 Data Visualization

Data visualization is representing the data by using visuals, to make sense, explore and communicate data. As the data that generated from social media increase over the years, the need of visualized data become required because visualizing this big data is easier than transferring them to spreadsheets. It is important for the analysts

that visualization enable them to notice the areas that have a problem or need to improve, clarify factors that affect customer behavior, and predict sales volume [6]. Although data visualizations have many different forms, but all of them have the major characteristics of order and production of information. Excellent data visualization should be: **Visually appealing:** the primary purpose of visual data appealing is to be a source of inspiration for the viewer. New technologies and tools for visualization increased effectiveness and usability. On the other hand, if data visualization applied with a poor visual appealing and with old tools and technology, it would be a waste of time and effort, because the traditional data visualizations tools on these days are unable to analysis.

Scalable: data visualizations scalability is the ability of information to handle the growth of the amount of data and users. Therefore, it is important to build visualization on a system that is scalable for accessibility and for future maintenance and modifications [7].

Accessible: accessible visualization allows modification when necessary and is easy to use; also, data must be accessible from any device and at any time.

Gives the user the right information: it is important before creating a visualization to identify exactly how it will be used, that helps to get the information that really needed [8].

3 RELATED WORK

Another use of data analysis is in the social media platform "Instagram" which can provide faster and easier communication. Instagram is one of the most well-known applications which users tend to share their personal moments. Several researches have been conducted on this area, one of these researches showed that most popular accounts that have highest views and followers are celebrities engaged in fashion and sports fields. The research comprised that numerous celebrities use Instagram to share their significant moments with their followers which helps them to be up to date with the celebrities' news and make them feel connected and closer to them. Furthermore, most of the celebrities tend to promote ads via such applications which costs little but has profitable outcomes. The data of the research has 100 of different categories, such as sports, celebrities, fashion, entertainment, and media. After entering the data into Rapid Miner and filtering the data, the program filtered category 1 and the result is 75 celebrities out of 100. After the data analysis, the main focus was useful attributes rather than the rest which are Iposts and ER on Hashtag through using aggregate operator. After the elimination of unuseful attributes, arrangement of data through

using sort operator was the next step, and this was to make either ascending or descending order for input data set. The majority of international singers use Instagram to share their music and their latest news for their followers and supporters; moreover, they have a huge impact and influence and considered as the most influential celebrities in the industry. The data proved this fact by projecting that most of the well-known celebrities on Instagram are from the music industry after using Filter Example operator [9].

There are other researches conducted about Snapchat platform. The first research aims to understand what content people send specifically, why they send it. Because the nature of the use of Snapchat is hard to observe and testify because of the ease of self-destructive of the content used in it. An online survey was conducted and aimed to inquire about the precise details of every content each user sent and received through Snapchat. The results showed that users mostly send 'selfies', and add text and sketches with the content they send. They use it frequently at home, and mainly for interaction with family members and friends; moreover, users use Snapchat as a simpler and more fun alternative to all other messaging applications. Results also showed a constant increase in the use of Snapchat from different segments around the globe [10].

The other article examines how brand reaches out to customers with their advertisement. The examining study uses a quantitative content analysis technique, it codes a specific number of snaps from different brands within different categories such as food & beverages, beauty, etc. the findings gives an understanding of how companies promote their brand through snapchat, due to some factors and basic concepts. Findings also show that snapchat had lack of usage among the brands for community generation through the engagement of dialogues [11].

4 METHODS

Many challenges may be faced when using social media data, because of its large volume, and ambiguous content. Due to that, it will be difficult and may be impossible to analyse around 100 million daily users' activity using traditional analysis techniques such as SQL databases and Excel. Therefore, this project uses advanced tools like Rapid miner and Tableau, to extract interesting knowledge from the data that collected from Snapchat and it helps companies reach specific audiences based on their online and real-world interests and behaviours. So, Snapchat's data collected for eight industries for data analytics:

- Sport
- Celebrities
- Entertainment
- Fashion
- Technology
- Fitness
- Travel
- News & politics

For each industry, we collected data about the top ten accounts according to the number of views and followers' interaction. After that, we are comparing the number of views in each industry, and the average ages of the users using Rapid miner program. Totally, we collected 8 industries with over 78 top accounts names.

5 FINDINGS AND DISCUSSION

5.1 Rapid Miner

Rapid Miner is a data science software platform that supports the machine learning process, such as data preparation, text & data mining, results visualization, and predictive analytics [12]. Our data analysed by using some of Rapid Miner's operators:

• Normalize

First, we started with normalize operator to normalize our data before work in cluster analysis that in fact works with distances, because if we do not normalize our data, we implicitly give a lot more weight when carrying out the cluster analysis to the variables that measured on a large scale. Normalization will also repel missing values that may influence our clustering results. Therefore, we provided our data to normalize operator and standard we used Z-transformation, which subtract the mean of each of the different variables divided by the standard deviation, and doing so, we got nice normalized variables that have a mean of 0 and standard deviation 1.

• Clustering

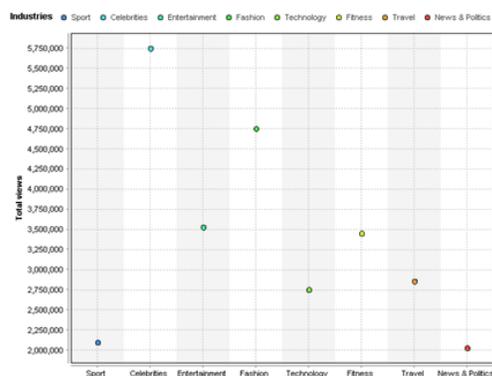


Fig. 1. The most viewed Snapchat industry.

We used this operator to define groups for the data. The approach used is K-Means algorithm which works in clustering data that have similar features or characteristics in one group. Since the K-Means algorithm works with numeric data only, so first we changed the type of data from non-numeric to a numeric type, then we identified value for K=8 (the initial centroid of the number of clusters should be detected). Figure 1 shows the results. The number of views in Celebrities is higher than any industry, while News & politics has the lowest views number.

• **Decision Tree**



Fig. 2. The most viewed on Snapchat at the decision tree.

Figure 2 shows the decision tree it is a popular technique in data mining, contain a collection of nodes that allows users to take problem with multiple possible solutions and display it in a simple understandable format. It is also called Classification Tree because it used for classification task. What we get from this operator that the total number of views of Sport industry is less than 2806500, and the views of Celebrities and Fitness industries are more than 2806500. Then it is obvious that Celebrities has more views than Fitness.

• **Predictive Analytics**

Age	2015	2016	2017	2018
0-11	0.6	0.9	1.2	1.4
12,17	11.3	13.5	14.8	15.5
18-24	16.6	19.3	20.9	22.1
25-34	11.3	16.7	19.1	21
35-44	3.7	5.4	6.6	7.5
45-54	1.7	3.5	4.5	5.3
55-64	0.7	1.8	2.4	2.8
65+	0.2	0.6	0.9	1.3
Total	46.2	61.7	70.4	77.1

Fig. 3. Snapchat users by age (millions) [13].

The statistic in Figure 3 shows the number of Snapchat's users by age for years 2015, 2016, 2017, 2018. To predict the number of users for the year 2019, 2020 and by using Rapid Miner tool, we used time series forecast and validation with ARIMA. In

Forecast Validation operator we set some of the data for the training window and some for the testing window. Inside Forecast Validation we insert the ARIMA model which trained on the training data set, for testing phase we added performance for testing data. The results in Figure 4 shows real values and forecasted values of 2019, 2020.

Row No.	forecast of ..	forecast of ..	Age	2015.0	2016.0	2017.0	2018.0
1	1.700	1.800	0-11	0.600	0.900	1.200	1.400
2	15.900	16.200	12,17	11.300	13.500	14.800	15.500
3	22.900	23.500	18-24	16.600	19.300	20.900	22.100
4	22.800	23.700	25-34	11.300	16.700	19.100	21
5	8.400	9.200	35-44	3.700	5.400	6.600	7.500
6	6	6.400	45-54	1.700	3.500	4.500	5.300
7	3.300	3.700	55-64	0.700	1.800	2.400	2.800
8	1.700	2	65+	0.200	0.600	0.900	1.300
9	82.700	86.500	Total	46.200	61.700	70.400	77.100

Fig. 4. The results of predictive analytics of Snapchat's users for year 2019, 2020.

Table 1: Snapchat users by age(millions) for years 2019,2020.

Age	2019	2020
0-11	1.7	1.8
12-17	15.9	16.2
18-24	22.9	23.5
25-34	22.8	23.7
35-44	8.4	9.2
45-54	6	6.4
55-64	3.3	3.7
65+	1.7	2
Total	82.7	86.6

5.1 **Tableau**

Tableau is a data visualization software that allows transforming data from different sources into dashboards, which helps in presenting data in a more meaningful and understandable way [14]. As found in Rapid Miner, that the number of views in Celebrities industry is higher than any industry; and the most viewed and popular category in Celebrities industry shows in Figure 5 after dividing the Celebrities industry into 7 categories in Tableau (Athletes, Media, Actors, Musicians, Models, Politics and Entrepreneurs).

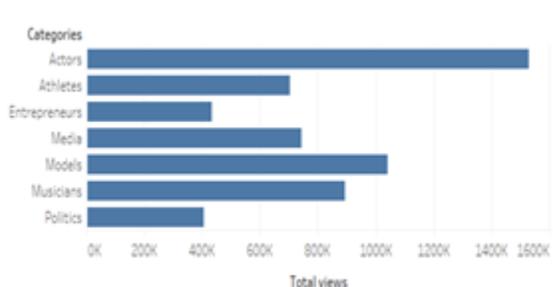


Fig. 5. Most popular celebrities' categories.

These information are valuable for companies, through which they can know the resonates content with users, the common behavior of their targets users, and to take the right decision regarding the appropriate industry for their advertisements.

6 CONCLUSION

This paper presented visual data analytics of Snapchat application, using advanced tools such as Rapid Miner and Tableau to analyze the big data produced by Snapchat and which cannot be analyzed by traditional analysis techniques. Production of knowledge and useful information hidden in these large data will help companies, especially in the marketing field to reach specific people based on their online behavior and to predefine audiences. The companies that use their data for analysis and mining to get valuable information having a competitive advantage than others, so we recommend the companies to benefit from their historical and big data in implement data analysis visualization. The results of our analysis show that the Snapchat's users interested in Celebrities industry more than other industries, specifically in the Actors category and that the number of Snapchat's users increasing by the years. In the future, we could extend our analysis to identify the average ages of the followers, and identify the visual culture and differences in content to help identify both the artistic trends and the trends of events in different places.

7 REFERENCES

[1] S. S. B. Shweta Patnaik1, "Social Media Analytics using Visualization," International Journal of Scientific Research Engineering & Technology, Bhubaneswar , 2018.

[2] Dummies, "What Makes Good Data Visualization?," 2018. [Online]. Available: <https://www.dummies.com/how-to/content/what-makes-good-datavisualizaiton.html>. . [Accessed 10 March 2019].

[3] K. Gordon, "statista," 2017. [Online]. Available: <https://www.statista.com/topics/2882/snapchat/>. [Accessed 10 March 2019].

[4] S. A. H. K. M. B. M. U. Samuel Fosso Wamba, "The Primer of Social Media Analytics," France, igi-global, 2016.

[5] P. B. J. B. Timothy Cribbin, "Doing social media analytics," SAGE Journals, UK, 2016.

[6] A. Singh, "Research Aids for Social Media Analytics," research gate, Andhra, 2017.

[7] S. D. Mico Yuk, "WHAT MAKES GOOD DATA VISUALIZATION?," [Online]. Available: <https://www.dummies.com/programming/big-data/big-data-visualization/what-makes-good-data-visualization/>. [Accessed 20 Mar 2019].

[8] "The Top Trends In Data Visualization for 2018," [Online]. Available: <https://carto.com/blog/top-trendsdata-visualization-2018/>. [Accessed 20 Mar 2019].

[9] A. M. R. H. a. G. F. Manar Saad1, "DATA MINING: Analysis of a Social Data Visualization Web Site (Instagram)," International Journal of Computer Networks and Communications Security , Saudi Arabia, Dammam , 2018.

[10] A. J. Lukasez piwek, ""What do they Snapchat about?," ScienceDirect, Bristol,UK, 2015.

[11] Y. A. K. B. L. M. U. d. E. P. R. Lina Gomez, ""Snapchat As An Influential Tool For Marketing Communication: An Exploratory Analysis Of Brands Usage"," Coronado, California, 2017.

[12] Rapidminer, rapidminer, 2019. [Online]. Available: <https://rapidminer.com/>. [Accessed 29 Mar 2019].

[13] Statista, "Percentage Snapchat users of January 2018, by age group," 2019. [Online]. Available: <https://www.statista.com/statistics/814300/snapchat-users-in-the-united-states-by-age/>. [Accessed 1 April 2019].

[14] tableau, "tableau," 2019. [Online]. Available: https://www.tableau.com/trial/tableau-software?utm_campaign_id=2017049&utm_campaign=Prospecting-CORE-ALL-ALL-ALL-ALL&utm_medium=Paid+Search&utm_source=Google+Search&utm_language=EN&utm_country=ME&kw=tableau&adgroup=CTX-Brand-Core-E-control&adused=32477696. [Accessed 1 April 2019].