Analysis of a Social Data Visualization Web Site: Twitter

Dalal Alwalmani[1], Nourah Alanazi[2], Maryam Alsadaan[3], Maryam Alomar[4], Mishaeal Aladaddh[5] and Reem Alhumaidan[6]

Department of Management Information Systems, Imam Abdurrahman Bin Faisal University (IAU),

Dammam, Kingdom of Saudi Arabia

ABSTRACT

Nowadays social media is a major intermediary to people communication whether for business or only entertainment. it is defiantly an effective way for exchanging different types of messages which of course influences people one way or another. Among many different social network websites and applications, we choose twitter. Twitter includes a huge number of users there for it contains a huge amount of data that are needed to be analyzed with the right tools to get the fuller picture of the data. For that purpose, we will present methods to accumulate the finding with an organized and easy to understand view using two of the business intelligence (BI) tools which are tableau and rapidminer that can be used by anyone whether an expert in analytics or a novice. In this paper we are giving a brief introduction on twitter and present some of the bigdata characteristics of a hashtags, related work to our subject and answer on the following questions, the purpose of choosing a hashtag to analyze, which hashtags we choose, why choosing them and what are the types of data used in the excel sheet. After that, we will show the result of the analysis used in both tools and demonstrate a collection of figures related to the findings. Then we concluded that data is not impossible to manage if the right tools were used.

Keywords: *Business Intelligence (BI), Tableau, Rapid Miner.*

1 INTRODUCTION

When talking about big influences and expressive web sites we are certainly talking about social media and one of these effective websites is twitter which contains bigdata that needs to be analysed according to user's needs whether personal or business. Twitter was created in 21st of March 2006, programmed by java, ruby, Scala, JavaScript. owned by Jack Dorsey when he first tweeted his first tweet at 9:50 pm saying that'' …we came across the world 'twitter', and it was just perfect. The definition was a short burst of inconsequential information, 'and' chirps from birds'. And that's exactly what the product was.'' and the headquarters of the company exists in San Francisco, California, U.S also they have more than 25 offices worldwide. This website is an effective method used to express ones' though through a brief content which contains 240 characters when it was only 140 characters which may include hyperlink. Also, a user can broadcast or post and attach pictures, videos and interact with massages known as ''tweets''. And whomever follows you, can see your tweets immediately. All of that is accomplishable only if you register in the website.one of the most important features of twitter is called #hashtags in which also known as trending topics such as #weekend and يوم_الجمعة# these two hashtags were the most used in the past year which any user can participate and comment on it. When you want to reach a specific tweet related to the topic just give it a click on the #hashtag and the list of tweets will pop up from a variety of users. This paper will discuss 4 of the hashtags, how to analyse them and represent them in two of the business intelligence tools which are RapidMiner and tableau. Tableau is a mining tool that represents visual analytics of the data that helps to see the story inside the data. Also, it is designed to do better decision in faster way. On the other hand, RapidMiner is a tool used to explore new ways of analysis to all kinds of data for the prediction of the future outcomes. [1][2]

2 RELATED WORK

In the past years there have been numbers of papers focus on hashtags of twitter such as On Recommending Hash tags in Twitter Networks paper (Springer, Berlin and Heidelberg 2012) they use a Twitter dataset generated by more than 150,000 Singapore users, one of the interesting findings is the significant use of new hashtags every day, which reflects most hashtags have very short life time. [5]

Also, the other paper is about sentiment analysis in twitter (Wang, Wei, Liu, Zhou, and Zhang 2011) they choose twitter because people express their feelings and opinions freely about various topics. They focus to find the different opinions on specific hashtags instead of reading from huge amount of tweets. [4]

3 METHODOLOGY

Twitter contains a huge number of data, and it is impossible to extract and analysis all twitter data such as images, videos, tweets and followers etc., so we focus in this paper on one aspect of twitter huge data which is Hashtag.

- *Why we choose hashtag data to analyze among other twitter data?*

There are many sites that provide information of a specific user account such as twitter analytics websites - https://analytics.twitter.com/ - and - https://foller.me/ - , but the limitation is that it requires authentication of the user account. However, there are some sites that facilitate extracting hashtag data from twitter in a specific period such as - http://hashtagify.me/ - and - https://ritetag.com - by writing the hashtag name in the search box, and it represent all data that related to the chosen hashtag such as the most country using it, percent of the gender, number of retweets and unique tweets.

Once the hashtag data is chosen, there are some obstacles to find prepared an excel sheet about a specific hastag, because the sites above doesn't provide excel sheet to be downloaded, so a manually fill up excel sheet which is the one used and analyzed it in this paper was the best solution.

- *What are hashtags we choose?*

There are huge number of hashtags in twitter, with different languages, thoughts, orientation and styles. This paper focused on a combination of hashtags in different two languages Arabic and English. The picked hashtag in this paper are(#weekend, #today, يوم_جمعة#, هلال#)

- *Why did we choose these hashtags? And what kinds of data that excel sheet contains?*

Choosing these kinds of hashtags did not come from futility, they consider as an active hashtags. The first hashtag was (جمعة#), we choose it, because it becomes trend in Arab countries every Friday. The second hashtag was (#weekend) it chooses,

because it's one of the most hashtag that all people around whole world use it and it active in most of the year. The third hashtag was (#today); people use this hashtag to express what happen to them during the day. The last hashtag was (هلال#) it's picked because it's the most popular football team in Saudi Arabia, and it has huge number of audience that they interact on hashtag all the time to support their team. All data of these hashtag are from last month which is March 2018 to make the analysis be up to date. Kinds of data inside excel sheet are (hashtag name, month, the most countries used, countries%, the most languages used, languages%, days, date, retweets, unique tweets , hashtag exposure). [6]

4 RESULT AND DISCUSSION

A. Tableau:

We use tableau to analyzing data that we collected to find the pattern inside the data, and as mentioned above tableau is used for data visualization. Here is some hashtag data visualization that we worked on:
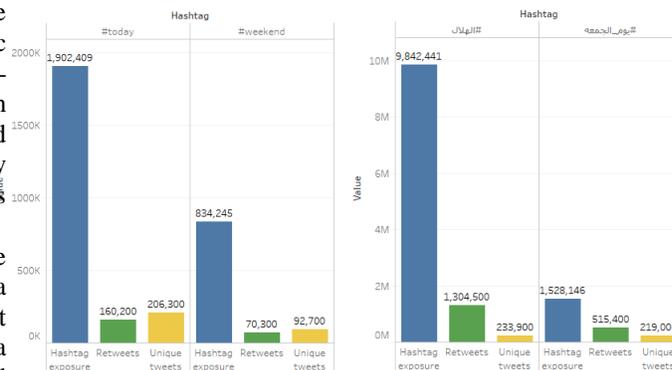
Fig.1. the total of (Hashtag exposure, retweets, tweets) for each hashtag.

This side_by_side bars represents the total of values inside excel sheet for all hashtags. We found that twitter users are mostly like to exposure and click on the hashtag rather than participating in it either by retweet the tweets that has been written before, or by writing a unique tweet. Moreover, we found that interaction is more in Arabic hashtags in our data. Also, for each hashtag there is a specific period or day that hashtag become more active. For example, the most active day for (يوم_الجمعه#) within the week is Friday, which means the number of tweets and retweeets are higher.

Fig. 2. the higher percentage of language and country.

This Packed bubbles represents percentage of the most language and country that used for each hashtag among other languages and countries.

B. Rapid Miner:

In RapidMiner we worked on a process sheet that contains 4 relations connected to 7 different operations that goes as follow:
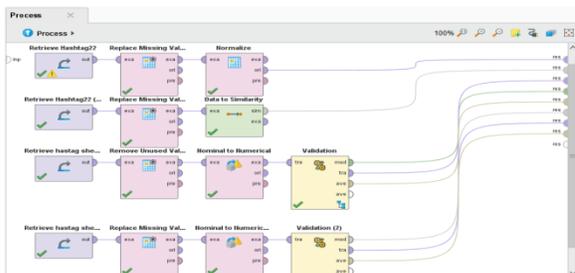
Fig. 3. All a Process

First, we normalized the data by retrieving inputs from an excel sheet and classified data according to the dates of tweets. And we 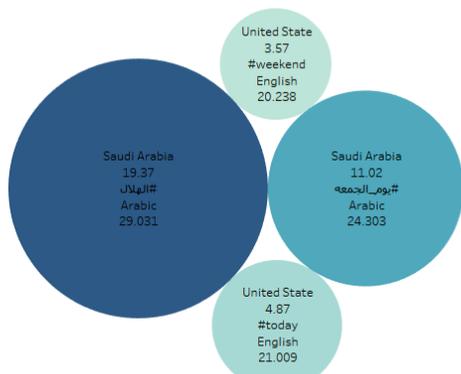operated replacement of the missing values 'white space' with the same value within the same margin using original port showing the primary data only 'positive values'. A bubble chart shows the wanted results accordingly with the days, hashtag and the kind of values was selected such as (Retweets, Unique tweets and Hashtag exposure).
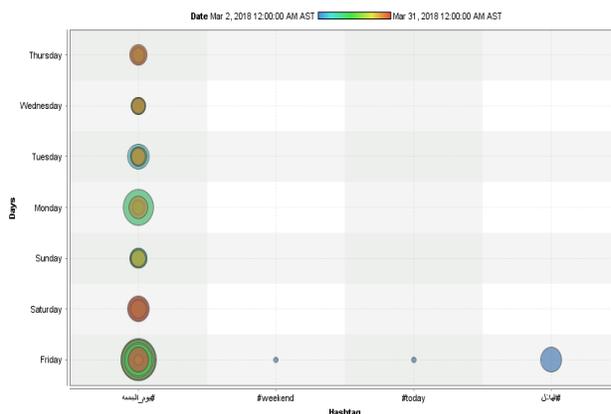
Fig. 4. Normalize data, original.

Second process was calculating the distance among similar dates tweets that are related to the hashtag which was demonstrated as Histogram layout using similarities meager port.
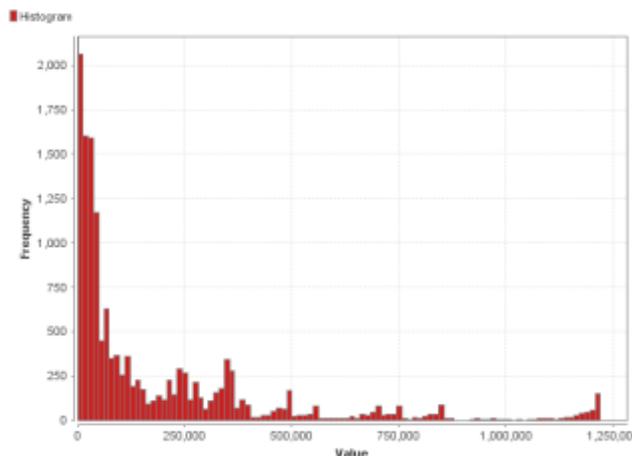
Fig. 5. Similarity measure object, Histogram, data to similarity.

Third process we retrieved the data from a second excel sheet according to the classified regions , accomplishing algorithms operation by removing unused values and take the output of this to covert it from nominal to numerical data and transfer it to the split validation operation to apply classification on the data, to make validation operation work we included 3 ports that are model ,training set and average, this operation is divided into training and testing, training includes decision tree then it will test the data through apply module connected with the performance 'classification' connecting it with unlabeled data port with apply model connecting it with performance through labeled data port. The results of this operation are shown as a tree (tight) chart contains regions and languages.

Fig. 6. Split Validation to use Decision Tree.

The last process we used replacement of the missing value operation connected to the conversion from nominal to numerical operation which is mandatory to cluster data through split validation with testing apply module connected to cluster count performance using 0.3 split ration with automatic sampling type and relative split parameters, as a result of this process we will get two root sets which are cluster0 and cluster1.

Fig. 7. Split Validation To Use Cluster Model.

For a gaining experience wise we used two values which are design and result to mine and present our findings according to our needed data but there was a third value that we did not use that is called auto model because it designs the data automatically and that is not what we needed

5   CONCLUSION

In conclusion, choosing twitter from various of social networking application was because, it contains big data which need to be analyzed. Moving the focus from different types of twitter data, toward one of the most effective data which is #hashtags. Moreover, for analyzing these kinds of data Rapid miner and Tableau are using which consider as analytic tools. A RapidMiner enables

users to create, deliver, and maintain predictive analytics for future outcomes. While, tableau is an analytic tool that used for data visualization. After analyzing twitter's #hashtag data we discovered that twitter users prefer to open and read about these #hashtags rather than interact with.

6   REFERENCES

[1] Best, B. and Thompson, M. (2018). - What is Twitter. [online] Tweeternet.com. Available at: http://tweeternet.com/ [Accessed 14 Apr. 2018].

[2] RapidMiner. (2018). Introduction to Advanced Analytics | RapidMiner. [online] Available at: https://rapidminer.com/resource/introduction-advanced-analytics/ [Accessed 14 Apr. 2018].

[3] YouTube. (2018). 1 مقدمة. [online] Available at: https://youtu.be/LfLqXJVPtqc [Accessed 14 Apr. 2018].

[4] Wang, x., wei, f., liu, x., zhou, m. and zhang, m. (2011). Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. [online] Available at: https://dl.acm.org/citation.cfm?id=2063726 [Accessed 14 Apr. 2018].

[5] Kywe, S., Hoang, T., Lim, E. and Zhu, F. (2012). On Recommending Hashtags in Twitter Networks. [online] Available at: https://link.springer.com/chapter/10.1007/978-3-642-35386-4_25 [Accessed 14 Apr. 2018].

[6] Ritetag.com. (2018). RiteTag |Trending Hashtags | RiteTag: Find the best hashtags. [online] Available at: https://ritetag.com/hashtag-search [Accessed 22 Mar. 2018].