Towards Improving Customer Behavior Prediction through Social Networks

Fayez Alrafeea, Moh'd Belal Al-Zoubi, Rana Yousef
The University of Jordan
Amman, Jordan
E-mail: fayez.alrafeea@gmail.com, mba@ju.edu.jo, rana.yousef@ju.edu.jo

Abstract

People use social media to create and share information, interests and ideas through the global network. In recent years, many social networks such as Facebook, Twitter, Instagram and others have become very famous. The use of social networks has occupied a significant part of the daily life of millions of people all around the world. In this paper, we demonstrate how social media content can be used to predict real world results. In particular, we build predictions about customers' satisfaction regarding different kinds of products after analyzing data collected from Twitter. We show how a model built from the confluence of tweets, twitterrers and polarity of text can be used to discover customers' satisfaction about some products. This in turn helps expect how sales figures will increase or decrease on the following months. Comparing our prediction results with real sales showed that a noticeable correspondence exists, provided that a suitable scenario and a good understanding of the business and the customer from the merchant side are taken into consideration to handle the specialty of products.

Keywords: social media, twitter, tweets, customer behavior, prediction

1. INTRODUCTION

Social media is a collection of online services that enables users to create and share contents and ideas in virtual communities and networks [1]. People use social media to chat and share interests, daily stories and events by interacting with each other using web based services enabled by social media.

The problem with social media content is that it's usually neither tapped nor categorized; this prevents users from making advanced searches, following interests or predicting people's trends. Accordingly, the content should be written in some structured way or using certain general criteria that have an impact on analysis, and can serve the prediction process. Twitter is one of the most famous social networking services that enable users to publish and read messages called "tweets". Twitter provides a huge data source of people's opinions and their trends about general issues, users tend to discuss their activities, status, opinions, ideas and interesting news stories through Twitter. Social media might be an effective means to examine trends and popularity in topics ranging from economic, social, and environmental to political issues.

The main goal of this study is to enhance the data analysis process of a social media network, specifically Twitter, to make predictions for people's opinions about a specific product, it will also show how the data collected from Twitter about products could be very useful for merchants to analyze and get useful feedback about their customers' trends and satisfaction in order to make plans and predict sales, it also shows the importance of the good understanding of customers' behavior on different times of product's lifecycle in order to get a useful results and build a correlation between customer's feedback on social media and the real behavior on ground.

In this study, we used data from Twitter to make analysis and build predictions about customers' satisfaction by processing and classifying data from Twitter that is related to three products (iPhone 6, Toyota Camry 2015 and "The Martian" movie). The results that will be achieved by applying our methodology will be compared to real sales' figures for the three products using different scenarios.

In section two we present related work to our research, section 3 describes our methodology to build our prediction model, sections 4 and 5 presents and analyzes the results. Finally, section 6 concludes the paper.

2. RELATED WORK

Much work has been done related to social media in general and to Twitter in specific. Simsek and Ozdemir [2] analyzed the correlation between stock market data and twitter data whose messages contain stock market words. Jin, et al. [3] tried to notify stock market investors by studying their emotions on social network in order to improve the prediction accuracy of the stock price. Liang and Dai [4] informed that tweets that have opinions are important because whenever people and organizations need to make a decision, they want to hear their customers' opinions. Lim and Buntine [5] stated that when customers make a purchase decision, a key factor of their decision can often be the reviews written by other consumers. Cvijikj and Michahelles [6] stated that public opinions in the form of trends are interesting not only for individuals but also for news reporters, pointing out to the fast-evolving news stories, sociologists, and marketing professionals. Boutet, et al. [7] gathered huge tweets data across one week for the 2010 UK general election and developed an algorithm to identify the political trends of users. Zhang and Skiena [8] used a news aggregation model along with the Internet Movie Database (IMDB) data to predict movie box-office numbers; they used quantitative news data generated by system for large-scale news analysis in order to predict movie grosses by analyzing two different models regression and k-nearest neighbor models.

Reviewing previous studies, we can see that most of this work is concerned with the sentiment analysis area where researchers tried to figure out the topic that people are talking about (topic detection) or the polarity of text (positive or negative), and then apply their model on some product predict sales' figures or elections to predict elections' result. However, few researchers build classification model based on the characteristics of the tweet itself (e.g. retweets' count) to build predictive model that can make such results. Although the sentiment analysis is out of scope of this study, we took it into consideration in our experiment because of its importance to determine the polarity of tweets' text. The other little papers used quantizing approach on different ways to estimate how much people are interested about a product and how the number of tweets could be compared with sales and used to detect income. However, researchers who adopt the quantizing model only take into consideration the number of tweets by counting them equally without giving importance to the effect that every tweet makes differently than the others which is realized very well in this study by classifying each tweet according to the effect it makes on people opinions. In addition, all previous studies focused on tweets only without stating the importance of the user (twitterer) which is recognized by this study because of its important role of making tweets more/less effective.

3. METHOD

The idea of this research is based on the opinions that people express through tweets. In addition to the people's thoughts, this small chunk –tweet- contains a lot of hidden information that can be used to evaluate this tweet and any bulk of tweets accordingly. When someone writes a tweet, he/she actually express their ideas about some public issue, event or product. If we have a huge amount of tweets about some product from a large number of people, we are actually able to get a very good feedback about that product. In addition, Twitter works like public media, that is; every tweet spreads the idea that the twiterrer thinks about, this idea sometimes makes a big influence on others' opinions while at other times not much people will care about it, the influence size depends on how much the twiterrer has popularity and influence on his/her followers. As a result, a tweet about some product not only express its writer's opinion, but also affects all its readers in somehow about that product and makes them more or less willing to buy this product.

Being able to estimate the influence that tweets make about a product during a period of time helps us to get an idea about customers' orientation and satisfaction about that product, which in turn could be used to detect how their behavior will be for the following months.

To estimate this influence, we took into consideration the three previous mentioned factors and combined them in different scenarios, then we analyzed combination results according to time, these factors are: tweets, twitterers and text polarity. In our study we combined these three factors in different ways and tested how this combination could be useful for merchants to analyze their customers' behavior.

3.1 Tweets

In order to discover each tweet's influence on people's opinions, we developed a regression model that assigns a value between 1 and 5 for each tweet according to how much they are convenience for people who read them, to accomplish this, we examined 11 factors (attributes) that we think are the most effective, these factors are listed in Table 1.

In order to estimate the influence that is made by each attribute, we developed a learning method that takes a group of tweets with their influence mark (between 1 and 5) as a test data and used them as entry to learn the real tweets data. A survey is used to give a mark for the test tweets by asking a group of people about their opinions about each tweet's influence as shown next.

Table 1 Tweet Attributes

Attribute	Description
Retweets count	When a tweet is retweeted for a number of times, it means that this tweet is popular and more people accept it.
Is truncated	If a tweet text is not complete, it's less possible to have popularity because truncated text usually means that the idea is not complete.
Contains media	If a tweet contains a photo or video, it's most likely be more convincing, media usually works as an evidence for the typed text.
Contains URL	URL is a link inside a tweet text, a lot of Twitter browsers usually display the link as a snapshot inside the tweet, when a user clicks on it, it will go directly to the linked page, and this gives this attribute the same evidence effect as "Contains media" attribute.
Contains other hashtags	Hashtag (#) represents the subject that a tweet talks about. When a tweet has many hashtags it means that it's related to many subjects and consequently will make it very general.
Is first hashtag	If a tweet contains many hashtags it will be less convincing. However, if the searched hashtag is mentioned as the first one, it means that it's the most related one.
Contains mention	Mention is when some user targets another by writing "@" before his/her name, mention means that this tweet is most likely related to a specific person and less related to others.

Favorites count	When a user reads a tweet and marks it as favorite it means that this tweet is more convincing for this user, more favorite counts means more popularity.
Is direct message	A direct message on Twitter is when a user mentions another user on a tweet beginning; usually means that it is related only to the mentioned user and this should decrease popularity.
Contains emotions (emojis)	Emotions or emojis are small icons added as a part of tweet; most likely these small icons make tweets more attractive and hence more popular.
Is retweet	If a tweet is a retweet from someone else, it will miss its original influence and hence reduces its effect.

3.1.1 Test Tweets

A group of 45 tweets about iPhone 6 with different criteria values have been chosen as an input for the survey, almost every tweet has different group of values to make sure that we have as much variety as possible. An online survey was prepared and sent to a random group of people, each person received a group of 15 tweets and asked to give a mark between 1 (less convincing) and 5 (most convincing) for each tweet about how much the opinion in this tweet is convincing for the person in case they want to buy an iPhone 6 and not deciding yet. The survey's results with their tweet attributes' values were used as a training data in the regression model in order to be able to classify the real tweets.

3.1.2 Real Tweets

Any product can be chosen in this research. However, the more talking about the chosen product on Twitter, the more precise the results will be.

To ensure that we have enough diversity, we chose three different types of products: mobile phones, cars, and movies. In addition to type diversity, we ensured to have time diversity as well, that is; the time cycle of sales which could be daily, monthly, or quarterly. According to these facts we have chosen the products mentioned in Table 2.

Reading tweets process spanned over several months to ensure getting enough number of tweets and to be able to read tweets continuously without time gaps. Some criteria were used within searching such as "from" and "to" to specify duration in addition to "language" which is always set to English "en".

As a result, more than 120 thousand tweets have been read for the three products as shown in Table 3.

Table 2 Product Types in our Study

Product Type	Product	Assessment Period
Mobile phones	iPhone 6	Quarterly
Cars	Toyota Camry 2015	Monthly
Movies	The Martian	Yearly

Table 3 Counts of Tweets for each Product

Hashtag	Duration	Count
#iphone6	15 Days	4812
#camry2015	10 Months	42,588
#TheMartian	12 Months	73,154

3.1.3 Processing and Saving

Before saving tweets to the database, we performed some processing to make them ready when needed to be read again for analysis. That is; we built a program that read every tweet, get its attributes, and change the format of some attributes, then save it to the database as a new record.

In addition to processing tweets' attributes, we had to process the user (twitterer) information for each tweet and classify users as will be shown later. Processing every tweet is performed after finishing reading from Twitter REST service to make sure that connection resources have been released before consuming any additional time. As a result of this stage, we have generated a group of tweets about each product saved to database and ready for later analysis.

3.2 Twitterers

The twitterer (user) attributes mentioned in Table 4 play a major role to influence his/her followers.

Table 4 Twitterer Attributes

Attribute	Description
Followers count	As the number of followers increases for an account, it has more impact and popularity.
Friends count	Friends count should have direct correlation with user impact but expected to be less than followers count.
Statuses counts	More statuses count means that a user has more followers and more popularity.
Is verified	Twitter gets users verified by checking their identities, this happens most likely for famous and well known users, when a user is verified then his/her tweets will be more trustworthy.
Listed count	When a user exists in others' lists, this means that more followers are interested about this user.
Profile image exists	The existence of a profile image for an account means that higher possibility that this account is active and not fake.
Banner image exists	The presence of such image gives more possibility that this account is active and is used usually.
URL exists	When a URL is linked with some account, it gives more possibility that this account is for a trusted party and makes their tweets more reliable.

3.2.1 Sample Twitterrers

In order to perform a classification for users to measure user impact on his/her followers, we choose a sample of 142 profiles taking into consideration the variety of all attributes (mentioned in Table 4) that could affect user's popularity. This will help us to know how each attribute makes the user more or less popular.

A class for each profile is set manually by examining the profile and figuring out its popularity and how much other users interact with tweets written by the user of the profile. The sample users have been divided into two groups; the first one is for training which contains 95 profiles while the other one is for testing and contains the remaining profiles (47), the classification model has been built programmatically using Weka implementation for decision tree (J48) [14], which produced an acceptable testing result, that is; from the 47 testing instances 9 (19 %) instances have been incorrectly classified while the other 38 (81%) have been correctly classified.

3.2.2 Real Data and Classification

As every tweet has a user that should be saved in the database to be classified, whenever we read a tweet we get its user and then add it to a list. As soon as we finished reading and processing tweets we loop over all users to extract their attributes then add them all to the database.

Having the model ready and data processed in the database, we have to pull instances and apply the classification model to have a class of one of five classes (VL, L, M, H, or VH) for each user. As a result, users for all tweets have been classified as in Table 5.

Table 5 Counts of Twitterrers Classes

Product	Users Count	VL	L	M	H	VH
iPhone6	31,812	2,814	20,113	8,771	89	25
Camry 2015	28,098	1,991	17,628	8,365	91	23
The Martian	1,562	138	912	448	52	12

3.3 Text Polarity

Text classification in this study is the process of knowing the polarity of text according to two types of classifications: subjective/objective polarity (SO Polarity) and positive/negative polarity (PN Polarity).

3.3.1 SO Polarity

Every sentence –including tweet text- could be subjective or objective information, subjective information is based on personal opinions, interpretations, points of view, emotions and judgment. It is often considered not appropriate for scenarios like news reporting or decision making in business or politics. Objective information; on the other hand, is fact-based, measurable and observable [9].

How SO Polarity affects influence? Twitter users usually log on Twitter to read what others are writing as personal views of public things rather than searching for facts which is normally searched for using search engines, for example if we have two tweets and the first one is saying "The new #iPhone will be available in gold, silver, and grey colors" while the other one is "#iPhone Just one week and I'm missing my old phone, didn't like this one", now when user reads these two tweet they will certainly be more interested about the second one which will influence their mind whether positively or negatively, but in case they want to know what colors will be available for the new iPhone they will google it or open Apple web site to know this kind of information.

Knowing this; it's important to classify tweet text as either objective or subjective in order to build better analysis and comparisons about this fact.

3.3.2 PN Polarity

For subjective tweets text, it's important to know whether it talks about product in positive or negative way to be able to discover users' trends and how much they are satisfied about that product.

3.3.3 Knowing SO Polarity and PN Polarity

To discover both SO Polarity and PN polarity, we've used third party service called Sentiment140 [10], this service makes sentiment classification for tweets using distant supervision (Go, et al., 2009). It can perform classification for a bulk of sentences by running online service that receive requests and return responses in JSON format. We chose this service to use after testing it for 2100 tweets with different polarities; these tweets were previously manually classified into 3 classes: objective, positive and negative. The received results were good enough to use this service, that is; we got 1724 (82%) tweets correctly classified.

To use this service effectively, we made a JSON based text file containing the text content from all tweets we want to classify, and we added a new ID for every tweet text in this file so that when a response return we can combine every classification result with its original tweet object.

The response is a JSON file containing every sentence with its ID and the classification for this sentence, the classification could be 0 for negative, 2 for neutral, and 4 for positive, note that objective classification result is neutral. The count for each polarity class is shown in Table 6.

Table 6 Counts of Polarity Classes

Product	Count	Objective	Positive	Negative
iPhone6	73,154	44,512	16,144	12,498
Camry 2015	42,588	23,965	11,753	6,870
The Martian	4,812	984	2,803	1,025

4. AGGREGATIONS AND RESULTS

Our final results are based on aggregating the information which has resulted from classifying tweets, classifying text to different polarities, and that which has resulted from classifying users for a time cycle. The time cycle is a day, month or year quarter, depending on the product under study.

To create a final result that represents a time cycle, we had to combine classification results for tweets, text, and users, which in turn offer different options.

4.1 Options for Scores

We can combine tweets' classification results in two ways related to score: score sum or score average.

4.1.1 Score Sum

Get the summation of all scores (classification results) for some period of time. For example, if we are studying a product on monthly basis (e.g. Camry 2015), we sum all scores in the same month so that every month will be represented by one score.

4.1.2 Score Average

Get the average of scores for some period of time, this option would make a small number of tweets represents the whole time period because it ignores the impact of large count of tweets that exist in this time period by taking only the average for them.

4.2 Options for Polarity

A choice has to be made regarding the tweet's text, that is, whether we want to take into consideration all tweets, subjective, or positive (or negative) tweets, this is more related on how merchants would use the feedback of their customers.

4.2.1 All tweets text

This is the normal case that reads all tweets without any conditions on text classification.

4.2.2 Subjective Tweets

A merchant could prefer to only get tweets that criticize their product ignoring all other spam and advertisements tweets; this is done by specifying the text class to subjective.

4.2.3 Positive Tweets

The merchant may like to get the positive tweets feedback to make expectations for the sales on a next period; we can do that by specifying the text class for read tweets. The same query can get the negative tweets by changing the "text_class" to 0.

4.3 Options for Combining Tweets and Users

In the previous sections we explained classifying tweets from one side and classifying users from the other side but we didn't combine between them yet to get one score representing a tweet and its user. To do so, we have three options: ignoring users, taking all users, or choosing specific class of users.

4.3.1 Ignoring User

A merchant could decide to get tweets analysis without their users; this could make sense if we suppose that the strength of user is already embedded inside the tweet itself, that is; when some user has a high impact on his/her followers the tweet itself will reflect that impact by having more values and more score.

4.3.2 Tweet Score Including User Class

Another view is that any tweet could be more influencing according to its user (twiterrer) but not all influences can be detected because most of followers are usually silent, that is; they don't make any action such as retweet or make a tweet as favorite, so it wouldn't affect tweet's score, this is why the user classification is important according to this view.

4.3.3 Tweets for User Class

The last option is to get tweets only for specific class; the reason is that merchant may only be concerned about high impact feedback, that is; the impact of tweets that are written by users whose impact classification are "High" or "Very High".

5. ANALYZING RESULTS

In order to analyze the obtained results, we are going analyze each product's data with different scenarios in order to explore how different analysis could make different results. Note that each scenario is built by choosing a set of the aggregation options described previously.

5.1 Scenarios

As there are many options to generate results (as discussed in the previous chapter), many scenarios could be generated and used, choosing the scenario and generating corresponding result is more related to business and how merchants want to see their customers' behavior in different times with different product life cycle stages. Every scenario differs from the others by the set of options considered when computing the result; every change in these options makes different results.

5.1.1 Popularity Scenario

This scenario shows the degree of awareness of people for some product, and how much they are interested in it, it intends to show whether a certain product is famous and well known or not. To measure this scenario result we calculate the summation of all tweets scores with all text classification (subjective and objective) and add the effect of user for each tweet.

5.1.2 Satisfaction Scenario

This scenario indicates the degree of satisfaction about a certain product according to what customers are talking about this product, either positively or negatively.

5.1.3 No Spam Scenario

The purpose of this scenario is to know the subjective opinions of customers and ignoring tweets that represent personal messages or advertisements; this is very useful for products that have a lot of other parties who try to target customers to other related products.

5.1.4 No User Popularity Scenario

This scenario ignores the popularity of users' popularity and only focuses on the tweets; this could be useful if merchants would like to eliminate the high influence of some users' popularity on the feedback results.

5.1.5 High Effective People Scenario

This scenario is the opposite of the previous one (No User Popularity); it concentrates on specific class of people according to their effectiveness, choosing of the class is related to merchant needs.

5.1.6 Positive Talk Scenario

If merchant would like to know if customers are satisfied about some product, this scenario is the most suitable one, it only takes into consideration tweets that talk positively about that product

Applying different scenario on different products we have generated the charts in figure 1.

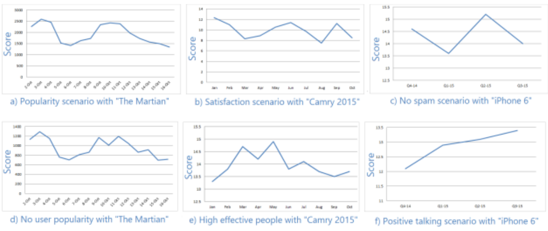

Figure 1 Scenarios applied on products

5.2 Comparing Our Results with Real Sales

The sample results presented in the previous section shows that our methodology could be useful to a great extent for merchants to build marketing plans, or making campaign. In addition, it can be used to detect sales behavior by selecting the appropriate scenario according to the merchant's business and experience.

When comparing our results to the real sales, we are actually comparing the changes in values of both figures, this means that when we have two values for the same period of time (e.g. February 2016) one on the real sales figure and the other on our resulting figure, we can compare trends of both values increasing or decreasing. Having the same trend means that we have a resemblance, while different trends means that we have dissimilarity, when applying this for the whole interval (e.g. one year), then we can know if our results are close enough to the real sales and accordingly if it can be used to predict future sales or assessing customer's satisfaction about a specific product.

5.2.1 The Martian

The chart in Figure 2 represents the gross profit for the first 15 days after launching "The Martian" movie.

If we compare this chart to the result of popularity scenario shown in Figure 1-a, we can see that there is a great resemblance between both for the same period of time.

As the popularity scenario represents the degree of which a product is famous without too much concern about its positivity, we conclude that it can be applied for simple products that don't cost a lot, this is because customers usually don't make deep search for this type of products and it won't be a big loose for them if it was not as expected.

For this type of products, the short time cycle for analysis could be very useful, daily cycle will show direct correlation between customers' feedback and customers' behavior on reality.

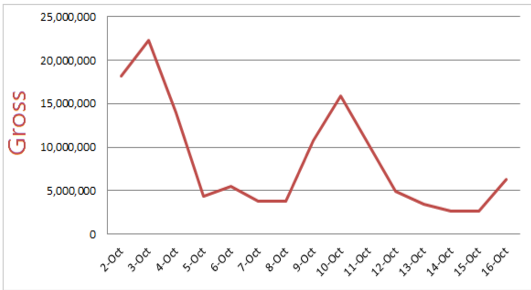

Figure 2 Gross for "The Martian" in first 15 days [11]

5.2.2 Camry 2015

This type of products requires analysis on monthly basis rather than daily basis, because customers take their time while searching and reading others' feedback before paying big amount of money to buy such products, the chart in Figure 3 shows Toyota Camry 2015 sales for the period Jan – Oct of 2015.

If we compare this chart with the satisfaction scenario chart shown in Figure 1-b, we can see that it's much closer to the chart of high effective people scenario that represents opinions from people with higher class in terms of convincing others, and this makes sense when we are talking about products such as cars.

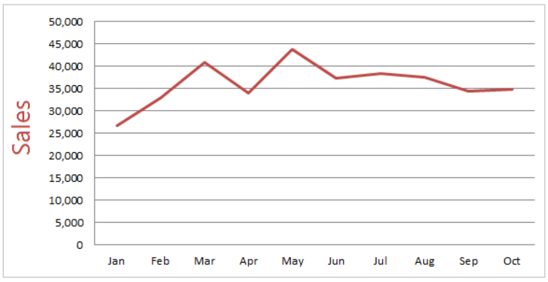

Figure 3 Camry 2015 sales during 2015 [12]

5.2.3 iPhone 6

This type of products is considered expensive but not as much as products like cars, so this product will take a place in the middle of the two previous products, that is; it's affected by the popularity tweets and propaganda that take a place in the first days of release while later buyers tend to get feedback from others and study it first before buying. The chart in Figure 4 shows iPhone 6 sales in quarterly base.

As the iPhone 6 sales started in September 2014, we can see that for the first months, the correlation is closer to the "No Spam" scenario (Figure 1-c) which represents popularity without objective tweets. However, in later months we can see that it is closer to the "Positive Talking" scenario (Figure 1-f), this result is expected, that is; for first months after release customers tend to buy it affected by propaganda and without getting deep feedback, while later buyers do need positive feedback from early users before buying it.

This leads us to the result that some products need to combine more than one scenario to understand the customer's behavior, if we combine the first two quarters from the no spam scenario with the second two quarters from the positive talking scenario we get result shown in Figure 5.

Combining the two scenarios ("No Spam" and "Positive") charts generates a new correlation that is close to the real sales chart.

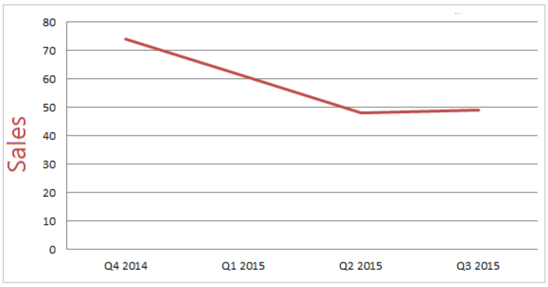

Figure 4 Sales for iPhone 6 since last quarter of 2014 [13]

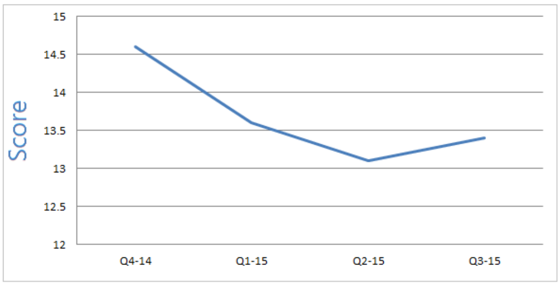

Figure 5 Result for combining "No Spam" and "Positive" scenarios for "iPhone 6"

6. CONCLUSION

In this study, we have shown how social media especially Twitter can be used to predict customers' behavior and get useful feedback about their satisfaction.

Since opinions on social media can be construed as collective feedbacks, we've studied its power to expect real world results. After conducting this study on Twitter data, we discovered that a content analysis related to a specific product can be used to make quantitative predictions that would express customer feelings regarding products in different times. In addition, it was found that this prediction could be used to make close expectations of how sales figures will behave during the following months.

We have explained the process of building a regression model for tweets and a classification model for users, starting from making a questionnaire to collect data then getting and preparing training data in order to assign specific values to tweets according to their impact on others. We also outlined the importance of twitterers by making classifications of users according to the influence they can make on their followers which in turn makes their tweets more or less effective.

The polarity of text had an important role in our experiments, which we have used to classify text into subjective or objective text, then the subjective text was classified into positive or negative, this classification of text provided an important role while applying different scenarios to predict customer's behavior.

Different scenarios related to tweets counts, relation with users, and polarity were presented, we discussed how these scenarios can be used by merchants to analyze customers' satisfaction and expect future behavior. In addition, this study was applied on different types of products (electronics, cars, and movies) to ensure the applicability of our method on any product type provided that we have enough information and good analysis.

Lastly, we've also shown the importance of proper analysis from the merchant side, which leads to good understanding of customers and prediction of the best time frames for issuing new models or making new campaigns.

7. ACKNOWLEDGMENT

This paper is based on a Master degree thesis at the University of Jordan by Fayez Alrafeea and supervised by Prof. Moh'd Belal Al-Zoubi and Dr. Rana Yousef.

8. REFERENCES

[1] Buettner, R. (2016), Getting a Job via Career-oriented Social Networking Sites: The Weakness of Ties. 49th Annual Hawaii International Conference on System Sciences. Kauai, Hawaii: IEEE. doi:10.13140/RG.2.1.3249.2241.

[2] Simsek, M.U. and Ozdemir, S. (2012), Analysis of the relation between Turkish twitter messages and stock market index.Application of Information and Communication Technologies (AICT), 6th International Conference on, pp.1-4, 17-19 Oct.

[3] Jin, X. Guo, D. and Liu, H. (2014), Enhanced stock prediction using social network and statistical model. Advanced Research and Technology in Industry Applications (WARTIA), IEEE Workshop on. pp.1199-1203, 29-30 Sept. 2014

[4] Liang, P. and Dai, B. (2013), Opinion Mining on Social Media Data. Mobile Data Management (MDM), IEEE 14th International Conference on, vol.2, no., pp.91-96, 3-6

[5] Lim, K. and Buntine, W. (2014), Twitter Opinion Topic Model: Extracting Product Opinions from Tweets by Leveraging Hashtags and Sentiment Lexicon. CIKM '14 Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management Pages 1319-1328

[6] Cvijikj, I. and Michahelles, F. (2011), Monitoring Trends on Facebook. Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing.

[7] Boutet, A. Kim, H. and Yoneki, E. (2012), What's in Twitter: I Know What Parties are Popular and Who You are Supporting Now! IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

[8] Zhang, W. and Skiena, S. (2009), Improving movie gross prediction through news analysis. WI-IAT '09 Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 01

[9] Su, Fangzhong, Markert, and Katja, (2008), From Words to Senses: a Case Study in Subjectivity Recognition. Proceedings of Coling, Manchester, UK.

[10] Sentiment140, Sentiment Analysis Service, (2016), http://help.sentiment140.com/.

[11] The Numbers, Research Service for Movies, (2016), http://www.the-numbers.com/movie/Martian-The#tab=box-office.

[12] Good Car Bad Car, cars information service, (2016), http://www.goodcarbadcar.net/2011/01/toyota-camry-sales-figures.html.

[13] Fortune, online magazine, (2015), http://fortune.com/2015/10/21/apple-iphone-q4-sales/

[14] J48, Weka API documentation, (2017), http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html.