Using Artificial Intelligence on Social Media’s User Generated Content for Disruptive Marketing Strategies in eCommerce

Marius GERU*, Angela Eliza MICU**, Alexandru CAPATINA***, Adrian MICU****

A R T I C L E    I N F O
Article history:
Accepted September 2018
Available online December 2018
JEL Classification
M10, M15
Keywords:
Machine learning, social media, clustering, Instagram, digital marketing, user generated content

A B S T R A C T
This study examines how user-generated content on social media can act as a marketing research instrument in identifying consumer behaviour. The analysis was performed on a sample of 900 Instagram images containing the hashtag #thegoodlife and labeled the digital photos using a machine learning algorithm. The results highlight the visual representation of the clusters resulted using the K-mean method and practical managerial implications.

© 2018 EAI. All rights reserved.

1. Introduction
In today’s digital-dominated business world, hard sell advertising and aggressive marketing tactics can act as a barrier in reaching the desired target audience. For this reason, to reach the hearts of consumers, one needs to find innovative ways to generate relevant and original content that encourages users into an open communication channel.

The leading candidate of open marketing communication is the hashtag, a novel digital instrument that creates online communities and invites consumers to join the discussion on brands and products. There is a new wave of social media platform that revolves around the hashtag philosophy and encourages users to engage in direct conversations with brands.

In this article, we will analyse how clustering images based on tags resulted using artificial intelligence can contribute to digital marketing strategies suited for eCommerce.

Starting with the consumer engagement importance in marketing we will follow how the concept of hashtags changes how brands communicate with a targeted audience in chapter one. Using the research methods described in section two, we will present the findings in chapter three concluding with managerial implication and further research that can be accomplished.

2. Theoretical Background
2.1. Consumer engagement
Mollen & Wilson (2010) states that there is a semantic conflict on how the academic world sees the online consumer engagement. They argue that the telepresence defined as the psychological state of ‘being there’ determines the online interaction. And marketers recount consumer engagement to sales, loyalty, and emotional investment.

The original source of consumer engagement is described by the theory addressing value co-creation with interactive experience and within marketing relationships (Brodie et al., 2011). Thus consumer engagement is seen as a multidimensional concept, depended by the marketing communication context, based on behaviour, cognition and emotion (Brodie et al., 2011; Hollebeek et al., 2014).

Stathopoulou et al. (2017) identify three type of brand engagement:

a. Cognitive - related to online consumer engagement that invests thought process in particular brand interaction
b. Emotional or affective - involves positive interaction that emerges bound with the brand
c. Behavioural or activation - energy, time and effort that consumer invest in a brand interaction

* *, **, *** Dunarea de Jos University of Galati, Romania. **Ovidius University of Constanta, Romania. Email address: marius@thecon.ro
Consumer engagement, particularly in social media relies heavily on hashtags as an instrument of marketing communication, as mentioned by Stathopoulou et al. (2017) is used in TV advertisements, printed media, radio and other mediums as well.

2.2 Social media, an endless source of data on consumer behaviour

The freedom to express personal information has reached new records, acknowledging the tools available in social media nowadays. Users are exploring unconventional ways to document actions, thoughts and opinions regarding particular interests on a diverse plethora of social media platforms. It is unreasonable to keep track of the information shared on these platforms (using traditional methods) because the content shared is diverse (text, images and video). Additionally users engage in interactions with other types of content like: shares, following specific brands and reviews. It is possible to collect and analyse content concerning commercial purposes. The concern for privacy is substituted with the need to establish a social status or express beliefs regarding certain aspects of public life. Maintaining this status and reputation on specific platforms has become a marathon and social obligation in some circles.

In eCommerce, the challenge is to transform this vast ocean of data into valuable insights, using predictive and prescriptive analytics in business research (D. Delen, H. Zolbanin, 2018). Profiling users before actually making the sale transaction and delivering relevant products and services to their needs is influencing the purchasing decision (R. Lixandroiu and C. Maican, 2015). The information can be found in multiple forms from different sources, and on most occasions, it is expressed as likes or ratings. However, gathered and evaluated can lead to highly relevant and engaging commercial offers in the digital marketplace.

The digitalisation of business processes in the last few years created an increasing demand for algorithms and processes that can interpret the vast amount of data. Often the research is concerned with consumer behaviour that already made the purchase and analyses data from existing eCommerce platforms. The scope of this article is to analyse social media behaviour and apply state of the art machine learning algorithms in order to find out how people interact and if particular patterns can lead to transactional marketing. The main focus will be on the concept of the hashtag and based on this notion we will do image analysis on Instagram using Google Cloud Vision API.

Social media has become the one stop shop for ideas, creativity, new products and lifestyle, is the corner piece to social and political moves and the real estate for trendsetters. There is a shift in the communication model, users using the high bandwidth of the mobile phones can now easily include photos, videos, gifs, maps, emoticons and other sorts of engaging content in conversations. For generation Z there isn’t an unreachable location, unseen place or unknown product, the question is how to present it from an unconventional perspective and find new meanings. Getting valuable and actionable data from this controversial user-generated content is a marketing challenge.

Relatively new social network service (SNS) Instagram got the attention of marketing researchers considering the fast growth it got lately. Many studies including computer vision have been published regarding the personality traits of users taking into consideration the pixels in images (Y. Kim, J.H. Kim, 2018). One interesting conclusion is in regards to food-related images that are significantly related to female profiles, this kind of information can have profound implications in marketing research and managerial implications. Abbar et al. (2015) use the rich data available on twitter posts in order to identify hospitality industry dietary habits and how social networks influence the behaviour of customers.

Even if most of the articles in social media is used in profiling users and understanding certain consumer behaviours Fatanti, M. N. & Suyadnya, I. W. (2015) discuss the benefits of user-generated content (UGC) on Instagram in promoting touristic destinations. Having visual endorsement from geolocated users, acts as a powerful incentive for prospects. The use of hashtags and curating quality images on dedicated accounts plays a massive role in the described marketing strategy. Brand identity and brand positioning are two critical concepts listed by Latiff et al. (2015) in a qualitative study on why do new business managers chose Instagram as a SNS. The simplicity of interaction with the content and the community that declines conventional approach to digital marketing makes Instagram a resourceful ground to young entrepreneurs. Employing advanced data mining and state of the art analytics with machine learning techniques (D. Delen, H. Zolbanin, 2018) can lead to unrivalled business opportunities.

2.3. Hashtags - brands invitation to open communication

Cappellini et al. (2018) addresses hashtag as powerful social tools that act as markers, enabling complex dynamics between actors in the diverse nature of the internet. Starting with the Arab spring, the concept of hashtags has increased its popularity addressing socio-political issues around the globe, and it became the de facto method in expressing a common point of view.

Gravitating on social issues is the default purpose of hashtags (X. Wang, et al.), also sporting events are covered (L.R. Smith, K.D. Smith, 2012), political debates (T.A. Small 2011), business research and digital marketing(B. Chae, 2015). A number of TV advertisement are placing hashtags on screen in order to create buzz, brand community and consumer engagement (A. Stathopoulou, et al. 2017). Instruments such as the hashtag used on social networks have led to changes in consumer habits. These habits are affected by new
ways of evaluating, searching, choosing and buying goods and services. These new tools influence how sellers work, their strategies, and how marketers are struggling with these challenges and difficult choices in the market (Thomas, 2007). According to Orzan et al. (2016), social media as a marketing communication channel has the ability to impact brand loyalty through two of its fundamental determinants: brand trust and brand affect. Communication between customers and firms helps on building brand loyalty beyond traditional methods. Hashtags drive to better product promotion, also allows brand followers to develop online open communities.

2.4 Artificial Intelligence - how to understand user generated content

Artificial Intelligence is an overemphasised term that requires, before any discussion some clearness. We can start by defining the term intelligence, which, according to Simmons and Chappell (1988), is an unconcealed ability to decipher problems and an inherent ability to learn escape procedures from problems. Taking this into account, we can determine artificial intelligence as a technology capable of finding solutions to a defined set of challenges. Understanding the environment in which this technology operates is an essential requirement. Then place into operation, the gathered data necessary to detect normal environmental parameters and recognises the functionality of the problem.

It can develop methods and conduct inductive reasoning to derive general principles learning from experience (M. Minsky 1961). The aim is to clarify how artificial intelligence can improve marketing methodologies and create models based on social media data that can be applied in e-commerce business strategies.

One of the compulsory artificial intelligence fields in marketing is computer vision. At a basic level, computer vision is a technology that recognises patterns, and more importantly, understands them. Both science and engineering need to collaborate in computer vision. (Ali Borji, 2017).

3. Research method

A good match for the tremendous volume of user-generated data in social media enables the emphasising of machine learning algorithms. Furthermore, taking into consideration the ever-changing nature of the content is a tiresome work to be labeled. For this operation, an unsupervised learning method like K-means algorithm can be used to find commonalities and cluster the information for better managerial decisions. Unsupervised learning is a subset of algorithms in artificial intelligence that does not necessitate any prior training and can be utilised for clustering and segmenting information. More specifically using the K-means method one can divide any n number of assertions into K clusters, calculating the means between elements iteratively. Used commonly in computer science as clustering this can be efficiently interpreted into marketing segmentation. A simplistic approach of the K-means clustering process described by Kyoung and Hyunchul (2008) is the division into four steps:

a. an initial number of clusters K is established and placed randomly on the graph assuming they are centroids
b. records are assigned to the nearest centroid forming k cluster
c. taking into consideration the clusters resulted in step (2), a new centroid is calculated using the mean distance from each record in the cluster.
d. Iterate Step (2) and (3) until the centroid position stops shifting or conditions are satisfied.

In this article, we have used K-means to do a marketing segmentation on the output of Google Cloud Vision of images on Instagram. The sample comprised 913 Instagram images and the relevant content (description, comment, likes, and location) attached within the hashtag #thegoodlife. The post-analytic qualitative process indicated that there is a tendency having a substantial commercial implication on this hashtag. The experiment can be replicated on any different hashtag on a SNS or set of images.

Technology being broadly available, an abundance of tools to analyse data becomes attainable on the market. Task’s that generally need intense processing power and advanced computer skill now can operate efficiently through a web application programming interface (API). Under this set of tools is Google Vision, a Computer vision (CV) technology that provides programmatic queries on specific images returning relevant tags containing objects and information found in the digital pictures.

4. Findings

The first step in this analysis was to collect 913 Instagram photographs with the hashtag #thegoodlife. In the following phase, Google Cloud Vision was used to label the images with up to ten tag keywords depending on the number of objects recognized in the image and what they expressed. The data was structured into JSON files including the tags resulted from Google Vision. Proceeding with the clustering using the K-means algorithm found in the sklearn python library, the experiment was focused a small number of clusters at the beginning (k=3), but the results that resulted were fuzzy. After extending the number of clusters up to 14, the segmentation was more precise and also the computation time was increasing exponentially.
Visual feedback of the data was obtained plotting the clusters number and with different colours. In order to achieve this, matplotlib, a python library used to generate charts.

Figure 1 shows that the segmentation of image tags is evident, for each of the 14 clusters was assigned an unique colour. Cluster 10 (in green) is differentiated away from the rest related to cuisine and hospitality industry. Also, there is a slight overlap between cluster 12 (tourist, hill, mountain landscape, nature), 13 (beach, coast, sea, yacht, bae), and 9 (hairstyle, long hair, chin, forehead). This overlap might be determined by users who taking pictures of themselves were located at certain touristic attractions. An interesting fact observed was that the algorithm classifies differently the images from seaside, mountains and images from other attractions.

![Fig. 1 Matplotlib output - clustering of instagram images](image1)

In order to visualise a particular cluster against all other data collected, another exploratory visual library for text analysis - Scattertext (J. S. Kessler, 2017). This library allowed the plotting of the tags that displayed the frequency of the tag as well.

![Fig. 2 Scattertext output - graphic and marketing materials](image2)

One noticeable result is the one that has marketing elements clustered together. We can see tags like text, font, logo, brand, design, product, graphics, graphic design. In figure 2 presented we can see the highest frequency of tags in the left upper corner, but there are other relevant ones, that are frequent in other segments as well (ex. product, design). It means that in other photographs available in the sample, these elements appear in the background or not as often as in the main cluster.
Considering the hashtag used to collect the posts was #thegoodlife it was expected that a cluster should form around tourism and vacations tags. In fact, there were 1556 relevant segmented tags out of more than 10000. This can be influenced by the fact that the data was collected in mid-September and Instagram is famous for travellers.

As stated by S. Abbar et al. (2015) cuisine is one of the main motives listed on social media, and one of the clusters resulting from the analysis is on this topic, see figure 4.

5. Discussion, implications, conclusions and future research

Images are becoming the preferred medium of communication on social media platforms, understanding the information shared by the users and applying this knowledge into marketing strategies can represent a competitive advantage. This analysis has implications both in understanding user behaviour but also for better targeting the desired audience. Understanding how different keywords cluster together marketing managers can create cross-selling campaigns or promoting the product including elements that usually are shown in images with the product.
Twitter ads has a feature that allows marketers to include specific keywords and it includes hashtags as well in order to target a specific audience. This option is not yet available on Instagram, but there is an option to target specific interest in Instagram and overlap this audience with a specific interest.

Product presentation in e-commerce has implication on user purchases intentions. J. Park et all. demonstrated how the image size and better online visual product presentation created a pleasurable shopping experience. Using computer vision, on analysing the content in the images can be taken one step further, allowing managers to identify cross-selling products, visible in figure 5.

![Fig. 5 Scattertext output - jewellery and accessory segmentation with K-means](image)

Social media is an untapped source of data using the right analysis techniques can lead to substantial managerial implications. Making use of this chaotic stream of information is the domain of machine learning. Even if there are a multitude of options available on the internet, there is no substantial academic writing outlining how they apply to the field of marketing.

Even if the plotting of clusters made the differentiation visible further empirical analysis is needed in order to confirm the findings. Future research can include the assessment of a statistical Chi-Square test of independence, to compare if there is a significant difference between the content of the hashtags (user-generated content) and the labelling done by a computer vision algorithm. In light of these insights, complex marketing applications can be developed taking into consideration dimensions like brands and products detected in images, geolocation in the metadata, sentiment analysis on description and comments.

Acknowledgement

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI – UEFISCDI, project number PN-III-P1-1.2-PCCDI-2017-0800 / 86PCCDI/2018 - FutureWeb, within PNCDI III.

References
6. Cappellini, O. Kravets, A. Reppel, Shouting on social media? A borderscapes perspective on a continuous hashtag, Technological Forecasting & Social Change


