Customer Profiling from Social Media Engagement using LDA and Sentiment Analysis Approach

Patrick Kanyi WAMUYU[1], Japheth MURSI[2]
[1]*United States International University-Africa, P.O. Box 1644-00200, Nairobi, Kenya*
Tel: +254 0722 309076, Email: kanyiwamuyu@yahoo.com
[2]*University of KwaZulu-Natal, Private Bag X01, Scottsville, 3209*
Tel: + 27 73 694 0723, Email: jmursi@yahoo.com

Abstract: Social media is now an essential component of the daily life of consumers. People usually share their interest, thoughts on brands and companies through discussions, tweets and status. At present, companies are competing to attract and meet customer needs. For companies, managing customer relationship through social media engagement has become a significant part of digital marketing strategies. The modern customer has different needs, expectations and behaviours which ought to be managed differently by companies., Customer engagement on social networks helps to create relationship with customers, and also acts as quick and cost-effective marketing tool. Social Customer Relationships Management (SCRM) provides a two-way communication channel between customers and businesses through social media sites. SCRM is based on a model of customer engagement which requires strong partnerships and interactions. The purpose of this research study was to understand customer interactions with business using topic modelling. The study analysed customer engagement on Twitter of Four selected banks in Kenya. We apply unsupervised topic modelling of LDA and sentiment analysis to create a profile of different customers of selected banks in Kenya. We focus on interactions from a consumer-centric perspective, not focusing on specific firm channels. We conclude that the extracted latent models not only provide insight to the consumer behaviour but also can also improve any company's Social Customer relationship management(sCRM) focused on different customer profiles.

Keywords: Social Media; Customers, Sentiment analysis; Topic modelling, LDA, CRM, sCRM

1. Introduction

The rise in adoption of social network sites in the last few years has changed the manner in which customers interact and exchange information. Social media is now an essential component of the daily life of consumers by enabling them to create and share content with fellow users and companies (Voorveld, van Noort, Muntinga, & Bronner, 2018) . For companies, managing customer relationship through social media engagement has become a significant part of digital marketing strategies (Malthouse, Haenlein, Skiera, Wege, & Zhang, 2013). With the fast growth and development of social network services, businesses have significantly improved their social media presence and commitments. Customer engagement on social networks helps to create relationship with customers, and also acts as quick and cost-effective marketing tool (Dolan, Conduit, Frethey-Bentham, Fahy, & Goodman, 2019). According to Chen (2017), customer engagement in social media serves as a crucial factor in influencing purchase intention among consumers and enhance the understanding of their needs and preferences, on the basis of shared information (Voorveld et al., 2018). While company representative is engaging with customers via social media, she/he is also marketing their brand to other users making social media a perfect customer relationship management platform (Wright, Nadler, Borders, Schwager,, & Sasnett, 2018).

With consumers actively expressing their views on products and seeking services online on social networks, social media has become a major source for social data mining (Litterio, Nantes, Larrosa, & Gómez, 2017). According to Forrester Research, organizations that are using insights from customer behaviour to drive their organizations will experience annual revenue increase of 27 percent between 2015 and 2020 (Hopkins , 2016). With data mining and machine learning (ML) techniques, businesses can learn from historical customer interactions and fine-tune the customer experience in a holistic manner by delivering the right content at the right time. Machine learning is far from replacing human engagement in the social media, however it increases the amount and quality of web interactions between companies and their consumers (An, Kwak, Jung, Salminen, & Jansen, 2018). Machine learning helps companies to build a holistic understanding of the customer through online engagement (Ransbotham & Kiron, 2018) . As companies attempt to optimize dialogue and develop engagement across multiple online platforms, ML tools help in analysing what type of content, keywords, and phrases are most relevant to desired audience (France & Ghose, 2019) . Businesses need to be responsive to customers' feedbacks and comments regarding their series and try at all means to respond to them.

This paper focuses on customers profiling segmentation based on their interactions with respective company's social media channels using of topic modelling attributes of Latent Dirichlet Allocation We hypothesise that using topic models and sentiment analysis would lead to better profiling and segmentation on social media interactions. Section IV presents actual experiment and implementation of these approaches on selected banks in Kenya.

2.0 Related works

2.1 Customer relationship management on social media

According to Peppers & Rogers (2011)Customer Relationship Management(CRM) comprises collection of customer information, segmentation of customers based on their expectation and needs, third is communication with customers through various channels and finally tailoring of products and service to these customers' practices includes activities like keeping in touch with customers, personalized messages, loyalty programs and other efforts to maintain profitable customer relations by satisfying customer needs. Social Customer Relationship Management (SCRM) focuses on interactions between both potential and current customers with organizations. The first stage in SCRM is to get to know consumers, their likes and dislikes (Paliouras & Siakas, 2017) Organizations ought to be responsive to customer feedback whether they are complains or positive comments. This helps recognize their underperforming areas and devise approaches for improvement. Its essentials to respond as non-action or response from the organizations can be inferred as a negative response by the customer.

CRM explores ways of treating consumers more as people by incorporating specific marketing and management approaches, such as customer loyalty and reward schemes (Tajvidi & Karami, 2017). Customers are important assets to a company and they each have different preferences, needs, expectations and behaviours so companies should be able to manage and treat them according to these profiles (Paliouras & Siakas, 2017). One of the ways of executing CRM is through social media data mining. Data mining is a series of processes to find relationships between items to form new knowledge patterns from large-sized data and processed with several techniques involving sciences including statistics, and mathematics. Through data mining we can extract information about customers that are important for effective strategy building (Marisa, Ahmad, Yusof, Hunaini, & Aziz, 2019).

Social Customer Relationships Management (SCRM) provides a two-way communication channel between customers and businesses through social media sites, such as Facebook and Twitter. Social CRM permits organizations to engage directly with customers in a simple manner while monitoring their interactions and influence as well as to track customer interactions and their social influence (Paliouras & Siakas, 2017). Social CRM has gained ground in modern businesses as a critical tactical resource for building and maintaining customer relationships. It gives companies the ability to maintain existing customers and obtain new ones (El Mehelmi & Sadek, 2019).

In today's business environment, businesses are faced with considerable operating challenges emanating from rapid technological advancement, extreme market competition, unpredictable changes in consumer' preferences and trends (Elena, 2016). Having satisfied customers is therefore the most important element for any organization that values its customer stakeholders. (Charoensukmongkol & Sasatanun, 2017) . Therefore, business ought to focus customer retention by managing customers' relationships effectively and trying to by understand their needs and expectations and able to tailor products and services to meet customers' expectations (Trainor, Andzulis, Rapp, & Agnihotri, 2014).

SCRM is based on a model of customer engagement which requires strong partnerships and interactions. Companies planning to use Social CRM must be conscious of the target group attributes and their degree of access to social media (El Mehelmi & Sadek, 2019). Engagement in the Social Media means that customers or stakeholders become active participants rather than viewers. The choice of suitable Social Media Networks is essential in achieving expected results according to the marketing strategy (Charoensukmongkol & Sasatanun, 2017).

The Stages of CRM Strategy

Social media engagement of a business provides benefits for its business like improved brand value (Hudson, Huang, Roth, & Madden, 2016) , customer trust and Stickiness (Zhang, Guo, Hu, & Liu, 2017) ; knowledge sharing (Munar & Jacobsen, 2014) and customer relationship management (Elena, 2016). Executing and building effective social customer relationship involves four stages (Sharp, 2003). These include:

i. **Interacting** – This is the first step of CRM where a contact is established between the customer and the business representative. This can be in form of inquiries on product information, service request and service complaints.
ii. **Analyzing** – The objective of stage is to analyse different customer information obtained from prior interactions with an aim of developing valued relationships.
iii. **Learning** – This stage entails learning and applying insights generated from analysis of customer interactions and behaviours to further strengthen the relationship with the organization.
iv. **Planning** - The final step is to prepare the best possible business strategy and to plan to meet the needs of each customer.

The above steps are essential in for implementing a successful customer relationship management in the present age of informed, demanding and educated customer on goods and services required (Paliouras & Siakas, 2017).

2.2 Customer profiling techniques

Customer profiling is the process of determining related characteristics or patterns of customer. Closely related to profiling is segmentation which is the task of dividing the information into small clusters or segments (Lanjewar & Yadav, 2013) . It is the use of data to identify or characterize a group of consumers or prospective customers. Segmentation is the classification of profiles into different sections. Approaches to profiling and segmentation can be divided into traditional and machine learning methods. The tradition method is the popular RFM model while machine learning methods include clustering and Topic models.

2.2.1 RFM: Recency, Frequency, Monetary Value.

RFM is one of the traditional customer profiling methods which originated from the catalog industry. It is an analytical model developed by Hughes (1994) used segmenting customers on their buying behaviour using three variables of Recency, Frequency and Monetary value. It has been used for improving efficiency of marketing efforts to existing customers (Scridon, 2008). RFM can be used to estimate the life value of customers.

Recency. This relates to the intervals of purchases by customers. It is classically the most powerful of the three variables for predicting response to a product subsequent offer. It posits that if a consumer has recently purchased product, he or she is more likely to come back and purchase same or a different product than an individual who has not bought anything recently (Mohammadian & Makhani, 2016).
Frequency. This represents the number of purchases in a particular period by a customer. This characteristic is second to recency in predictive power for response. The bigger the frequency the better.
Monetary value refers to the value of all purchase's money wise. This element is the least effective of the three when it comes to predicting responses but helps in better understanding. These three elements can be used alone or in combination with other features to aid in the profiling and segmentation of customers.

2.2.2 Machine learning methods

2.2.2.1 Latent Dirichlet Allocation (LDA)

This is a generative probabilistic model in machine / deep learning applications related to natural language processing. LDA is a generative unsupervised topic model used to capture semantic information from a collection of documents (Blei, Ng, Jordan, & Lafferty, 2003). It generates topics based on word frequency from a set of documents. It is a probabilistic model where documents are represented as a combination of the probabilities of belonging to each topic (Carrera-Trejo, Sidorov, Miranda-Jiménez, Moreno Ibarra, & Cadena Martínez, 2015). It takes documents as input and finds topics as output and generates the approximate percentage each document talks about each topic Using associated metrics, LDA can be used to classify the semantic relationship between a group's words. LDA treats each document as a combination of topics, and each topic as a combination of words (Silge & Robinson, 2017). LDA builds a topic per document model and words per topic model, modelled as Dirichlet distributions. Tirunillai and Tellis (2014) used LDA to extract latent dimensions of consumer satisfaction using consumers' online product reviews. The LDA is built on two basic assertions; documents that have similar words typically have the same topic and documents that often occur together have the similar topic (Malik, 2019). According to Büschken and Allenby (2016), LDA models lead to better inferences and forecasts of product ratings compared to simple word-based models. Blei and Lafferty (2007) illustrated the usefulness of subject models in modelling semantic text structure and drawing latent text patterns.

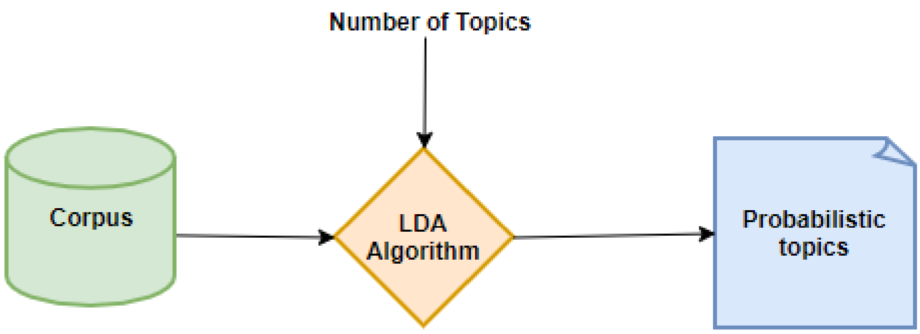


Figure 1: LDA Workflow modified from Carrera-Trejo et al., (2015)

2.2.2.2 K-Means

Another algorithm applied in profiling and segmentation is clustering using K-Means. K-Means unsupervised learning technique that is used in solving clustering problems. K-means algorithm divides n observations into k clusters whereby each observation belongs to the group with the nearest means acting as a cluster model (Coates & Ng, 2012). K-means' aims is to group similar data points and find underlying patterns. K-means searches for a fixed number (k) of clusters in a set of data to achieve this goal. A cluster refers to a set of aggregated data points due to some similarities. A target number of K is determined, i.e. number of centroids needed in a dataset. A cluster refers to a collection of data points aggregated together because of certain similarities. A target number of K I.e. number of centroids needed in a dataset is defined (Haraty, Dimishkieh, & Masud, 2015). Each data point is assigned to each cluster by reducing the sum of squares in the cluster.

K-Means Clustering was applied to build customer segmentation for targeted customer services using their behaviours and demographic profiles (Ezenkwu, Oloyede, Umana, Jerome, & Ekpo, 2013). This paper elaborates upon the use of the data mining technique of LDA and clustering technique to segment customer profiles for selected Banks in Kenya. Clustering can help identify trends and habits of customers purchase behaviours, boost customer service for better customer loyalty and therefore retention, (Lanjewar & Yadav, 2013) 2013).

2.3 Customer profiling and segmentation on social media

Due to the affordability of internet services to household users, online shopping has witnessed an unprecedented increase in recent years (Pater, Vári-Kakas, Poszet, & Pintea, 2019). Mining user interests from user behavioural data is essential for social media marketing and online advertising. Based on user interests, companies can significantly reduce service delivery costs by tailoring relevant products to their customer. Businesses ought to figure out about prospective customers ' preferences and build business models for those groups of customers. Knowing customer's personal values is therefore important. For a long time, the banking industry relied on traditional market segmentation techniques to tailor its services to customers (Paker & Vural, 2016) , and truly understand the needs of customers and provide that customer with the experience that best suits their needs. Traditional forms of customer profiling entails demographic, geographical, socioeconomic and psychographic characteristics (Dolnicar, 2007).

In addition, accurate consumer profiling supports companies ' marketing activities by encouraging marketing teams to focus on creating new messages and delivering them effectively through the most appropriate channels (Mohammadian & Makhani, 2016). Coupled with the provision of quality service delivery, segmentation enables positive word-of-mouth to be obtained and circulated among each segment's members (El Mehelmi & Sadek, 2019) Instead of examining the entire consumer base as a whole, segmenting consumers into clusters is effective (Wind & Bell, 2008). It helps recognize the characteristics of each group of consumers, and engaging each group with specific, more oriented marketing campaigns rather than just segmenting consumer demographic or geographic variables (Scridon, 2008).

According to Jadczaková (2013) segmentation as “a set of variables or characteristics used to assign potential customers to homogeneous groups”. Segmentation is important because a times some organization often limited resources and ought to concentrate on how best to classify and satisfy its clients (Jadczaková, 2013). Through segmenting, companies can recognize niche markets with particular needs, emerging markets to find new consumers and aid in delivering marketing messages that are more oriented and efficient (Büschken & Allenby, 2016). Customer engagement on social media is any online activity a customer makes in relation to business. This can range from inquiry about a service, comment on a product, likes, shares, purchases or subscription. Generally, customer engagement on social media is regarded as value addition from a customer to a business (Pansari & Kumar, 2017).

In the current digital divide, the importance of customer interaction goes beyond monetary prospect. It promotes the business services and products while strengthening customer loyalty and satisfaction (Pansari & Kumar, 2017). Social network sites when used effectively can help businesses to understand customer needs and motivates them to buy the products (Tajudeen, Jaafar, & Ainin, 2018) . Various research has been carried out on analysing user interest on social media. These include using analytics to understand how personality traits influence information sharing behaviour on social media (Deng, Lin, Liu, Chen, & Li, 2017) using topic tags to predict user brand preference prediction of community media (Yang, Pan, Mahmud, Yang, & Srinivasan, 2015) and using social media data to predict consumption behaviour (Zhang & Pennacchiotti, 2013).

This study reviews past research on profiling and segmentation in order to comprehend suitable approaches to customer interactions on social media. We apply topic modelling algorithms to create a profile of different customer of selected banks in Kenya. We focus on interactions from a consumer-centric perspective, not focusing on specific firm channels

2.4. Related works on customer segmentation and profiling

Ezenkwu, Oloyede, Umana, Jerome, and Ekpo, (2015), developed a k-Means algorithm and trained using a structured two-function data set of 100 retail learning data. The attributes were the average amount of products purchased each month by the company and the average number of customer visits per month. From the dataset, four customer clusters or segments were identified with 95% accuracy, and they were labelled.

Mohammadian & Makhani (2016) presented an overview of consumer segmentation based on customer lifetime value (CLV) and RFM process. They examined how organizations can collect customer data and analyze these data in order to obtain useful insights into the customer. The study proposed a new three-step approach using RFM analysis to categorize customers into categories that are rationally distinct. Using Customer Relationship Management, Zahrotun (2017) used digital customer data to determine the best consumer. Christy, Umamakeswari, Priyatharsini, and Neyaa (2018) explored the application of CRM for online shopping and demosnatrated that segmenting potetial customers could help organization boost their sales

A study by Lee and Bae (2016) sought to understand social media users by segmentation analysis. They segmented users based on their effectiveness on social media marketing and usage. Modern consumers have adapted quickly social networking sites, making SNS the most important developments in consumer-based IT (Lee & Bae, 2016). User-generated content can influence the willingness of customers to pay in the form of reviews. Hamilton et al., (2014) found that the presence of referenced markers in user-generated content improved the market demand for a product. (Wu and Wu (2016) asserted that individual willingness to pay varies according to personal values on uncertainty.

Wong and Wei (2019) sought to develop an online behaviour analysis tool and predict next purchases from an online air travel corporation. Their methodology included data mining, customer segmentation and predictive analysis of competitors ' pricing they identified and segmented customers based on their purchasing behaviour. Consumer understanding and segmentation are essential in order to retain high-value customers and to acquire new customers in any industry. Analysing online shopping behaviour, comments and interactions of segmented customers can provide useful insight into tailoring of customised services and products to high-value customers at the right time (Wong & Wei, 2018).

3.0 Methods

3.1 Data collection

Tweets were extracted from Twitter using Twitter API and Twitter developer account. The twitter data was crawled using rtweet library in R to access the Twitter API (Kearney, 2016). We used Twitter streaming API to extract tweets from selected banks. Not only does Twitter draw a diverse audience, it also makes it much easier to recognize the target audience. Characteristics and preferences can be easily discerned from profiles, tweets, retweets and hashtags (Mueller, 2019)

3.2 Data Pre-processing

In many text mining, data pre-processing is one of the key components in generating a useful topic model. Unstructured data by its nature is noisy, full of typing errors, bad grammar, usage of slang, presence of unwanted content like URLs and expressions (Pereira, 2017). The main objective of data cleaning is to identify and remove errors and anomalies in order to improve the quality of data for analytics and decision making. In this step data cleaning and pre-processing of the raw data is undertaken (Corrales, Ledezma, & Corrales, 2018). In general data cleaning involves eliminating special characters, and numbers, punctuations, words that are unnecessary.

3.3 Tokenization and Lemmatization

After pre-processing the corpus is split into segments blocks in preparation for feature identification and information analysis (Allahyari, et al., 2017). Tokenization is a process of splitting sentences into small parts called ‘tokens. In this paper tokenization is done using TweetTokenizer a python library for splitting texts to tokens. After tokenization, lemmatization is performed on the extracted tokens . Lemmatization uses vocabulary and morphological analysis of word and tries to remove inflectional endings, thereby returning words to their dictionary form (Balakrishnan & Lloyd-Yemoh, 2014).

3.4. Profiling and segmentation

The final task in the methodology is the analysis of sentiments and the modeling of topics. Sentiment analysis is the process of extracting information from people's opinions, assessments and emotions about occurrences as depicted in text (Im, Park, Kim, & Park, 2019). In this paper we sought to identify customer interactions that had positive, negative or neutral sentiments. This analysis was performed using the NLTK library (Bird, Klein, & Loper, 2009) sentiment analysis entails determining the polarity of text which can be negative positive or neutral. Text blob python library was used in this study. This paper sought to collect social media interactions from selected banks in Kenya and applied latent Dirichlet Algorithm (LDA) to perform topic analysis and derived behavioural characteristics of different bank users. Sentiment analysis of the interactions using Texblob Python library was also performed. Topic models are unsupervised machine learning algorithms which unearth hidden thematic structure in collection of documents. We apply LDA to 4 datasets of selected bank user interactions extracted from Twitter (Sokolova, et al., 2016). We hypothesised that, through LDA and sentiment analysis, companies infer latent factors beyond customer engagement with customer’s which can be used to create individual marketing material to existing customers. Latent Dirichlet Allocation (LDA) is a topic modelling algorithm used for extracting topics from a given collection of documents (Blei, Ng, Jordan, & Lafferty, 2003). Sentiment analysis is the study of opinions, feelings, and emotions expressed through writing We used Twitter’s API to extract tweets from 4 selected banks.

In this research, we were interested in finding out the tweets related to selected banks that had positive, negative or neutral sentiments. In addition, we sought to profile, and segment customers based on their interactions and comments on the bank’s Twitter pages. Below workflow shows our approach of segmentation and profiling.

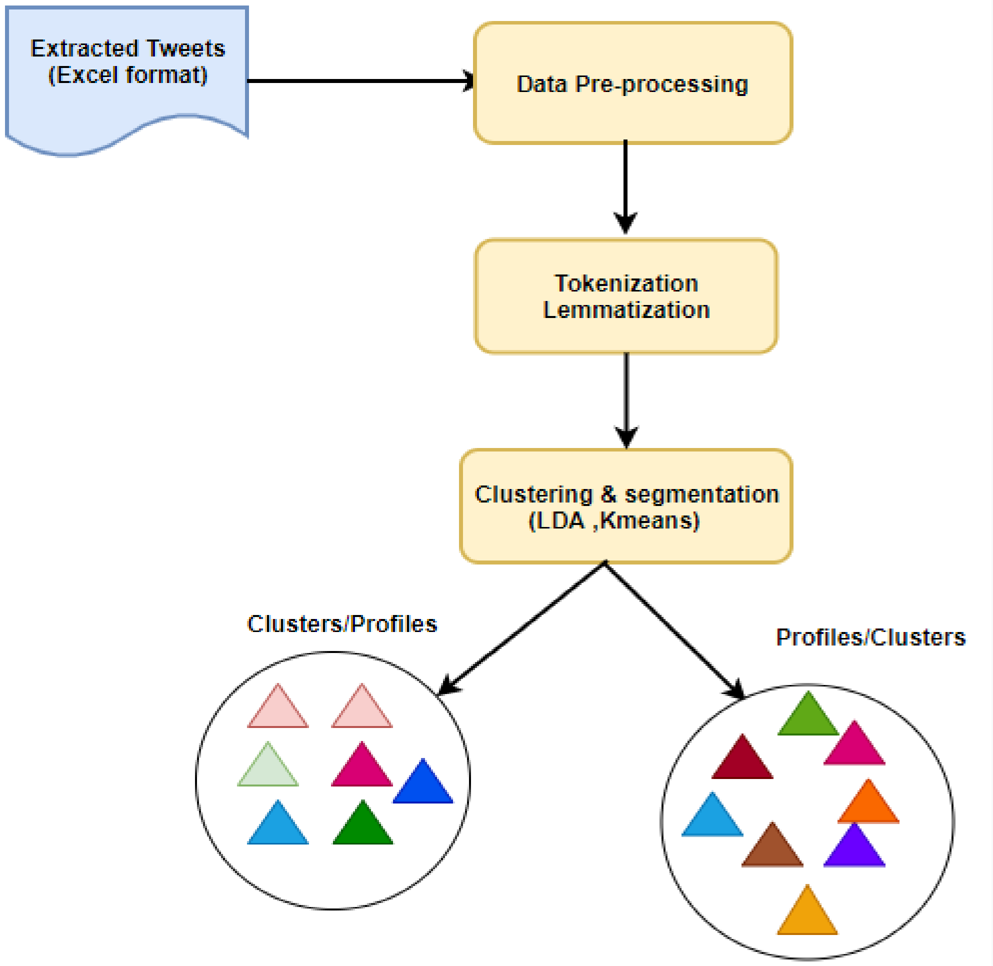


Figure 2: Segmentation workflow modified from (Islam, 2019)

4.0 Experiment and results

The results show that LDA can be used for clustering of social media engagement between customers and clients improves topics interpretably. Embedding of topics also helps addressing a classification problem.

In addition, the results of the study showed that there was a significant difference in how different companies respond and address their customer needs based on the selected banks. There were instances where several interactions were flagged negative based on how customer was responding and voicing their opinions on service delivery. Negative sentiments had delayed responses from the bank, denial of services and inconveniences in banking caused as dominant topics.

We chose four large banks in Kenya for the purposes of this study. The banks names have been changed to protect their identities.

4.1: Top words per topic and Topic label

Table 1: Topic distributions

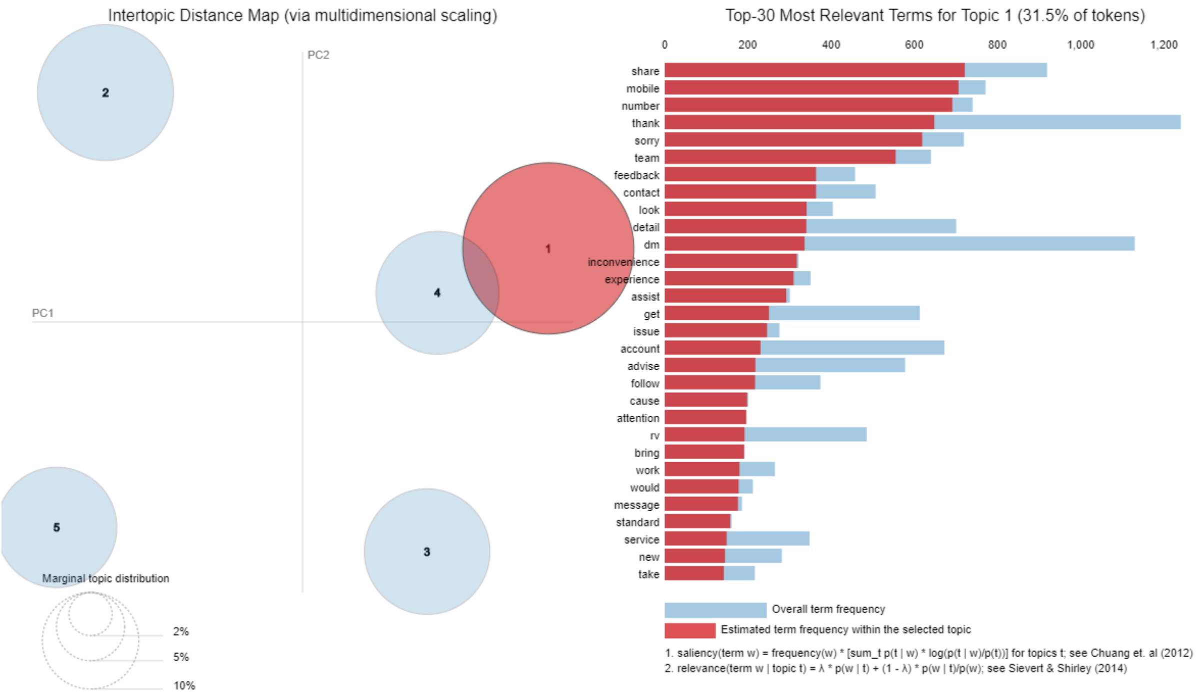
Top 10 words per topic with weights as Identified by	Topi	Inferred Topic Label

LDA for four banks	c	
(1, 0, '0.020*"inconvenience" + 0.019*"branch" + 0.018*"account" + 0.017*"team" + ' '0.013*"follow" + 0.013*"app" + 0.011* + 0.011*"know" + 0.011*"good" + ' '0.010*"")	1	Customer complaint
(2, '0.059*"respond" + 0.028*"thank" + 0.021*"bank" + 0.019*"check" + 0.016*"client" ' '+ 0.013*"year" + 0.009*" welcome" + 0.009*"africa" + 0.008*"shortly" + ' '0.007*"code"')	2	Feedback
(3' '0.032*"share" + 0.031*"mobile" + 0.031*"number" + 0.029*"thank" + ' '0.028*"sorry" + 0.025*"team" + 0.016*"contact" + 0.016*"feedback" + ' '0.015*"look" + 0.015*"detail"")	3	Assistance of Banking/card issue
(4, '0.019*"detail" + 0.014*"run" + 0.012*"share" + 0.011*"thank" + ' '0.011*"follow" + 0.010*"loan" + 0.010*"access" + 0.010*"personal" + ' '0.010*"account" + 0.009*"select"'),	4	Loan
(5, '0.061*"dm" + 0.034*"check" + 0.028*"touch" + 0.024*"see" + 0.021*"advise" + ' '0.020*"get" + 0.014*"card" + 0.012*"response" + 0.011*"detail" + ' '0.011*"thanks"')]	5	Feedback

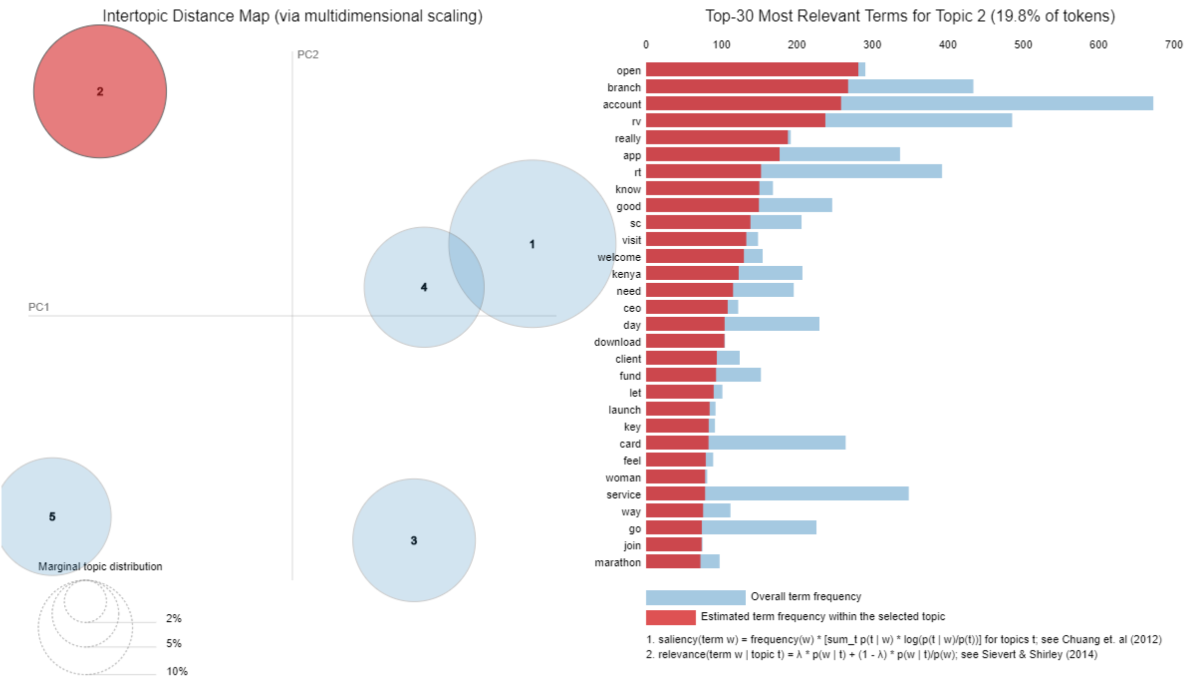
4.2 Visualization

We use Sievert and Shirley's LDAvis (2014) to visualize our LDA latent inferred topics Gensim's pyLDAVis is the most widely used method for visualizing the data in the topic models (Sievert & Shirley, 2014). In LDAvis, each bubble represents a topic and in this paper 5 topics were generated from the dataset as shown below.

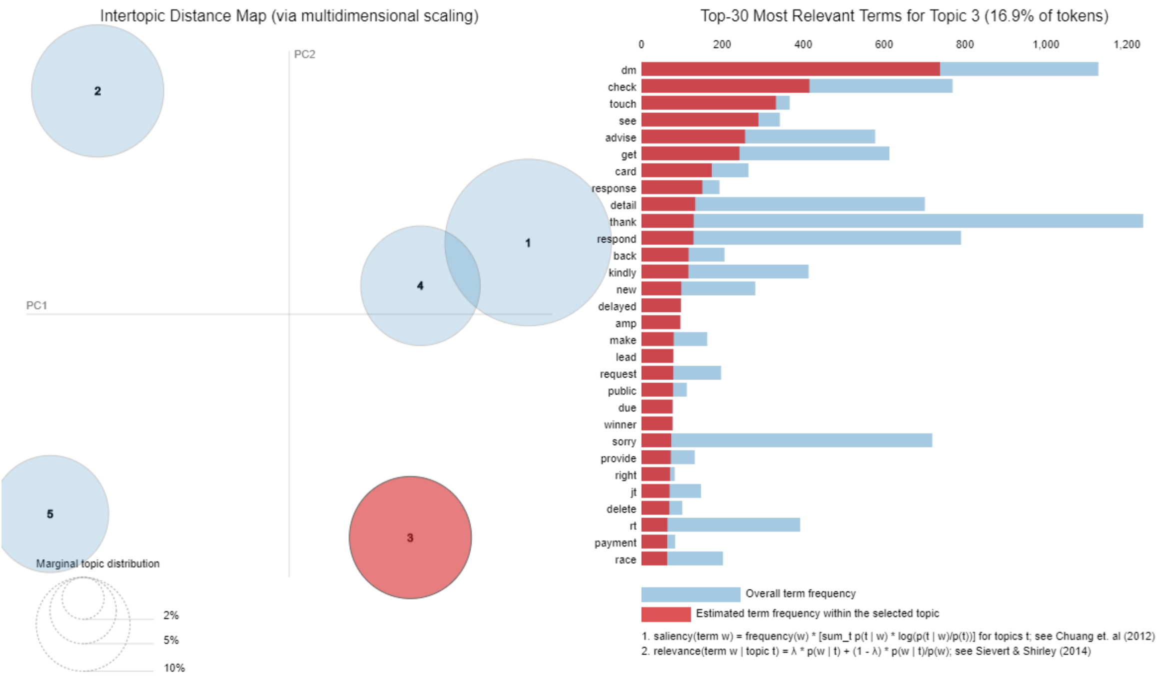
4.2.1: Topic 1



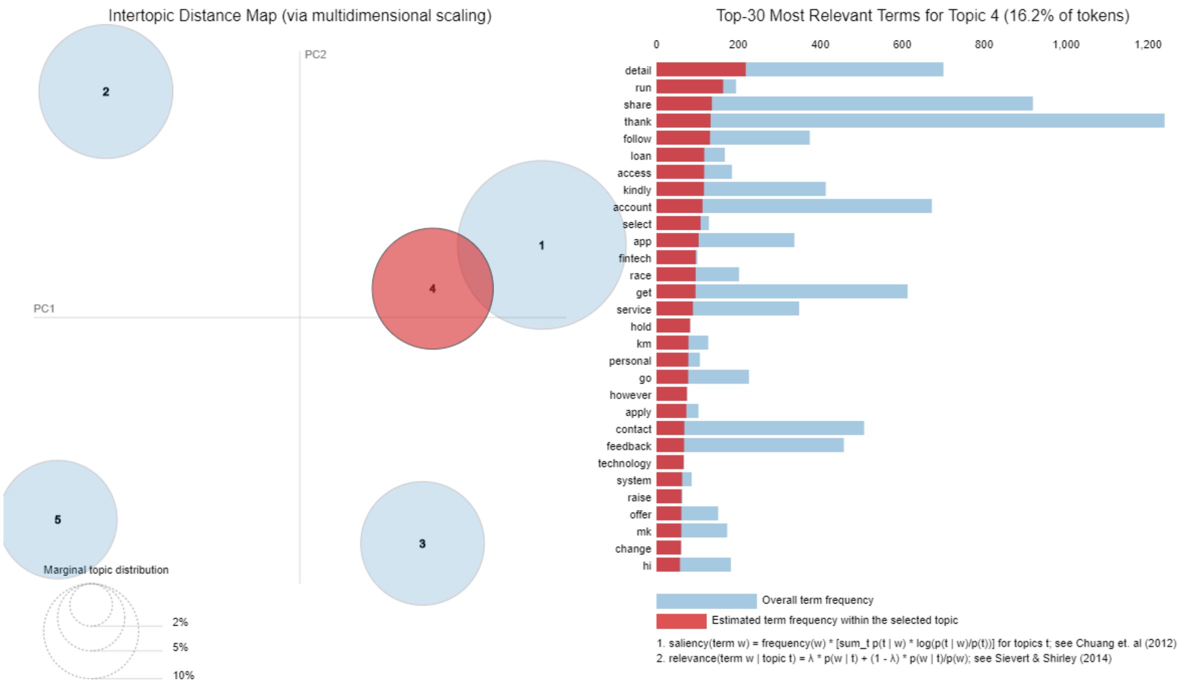
4.2.2 Topic 2



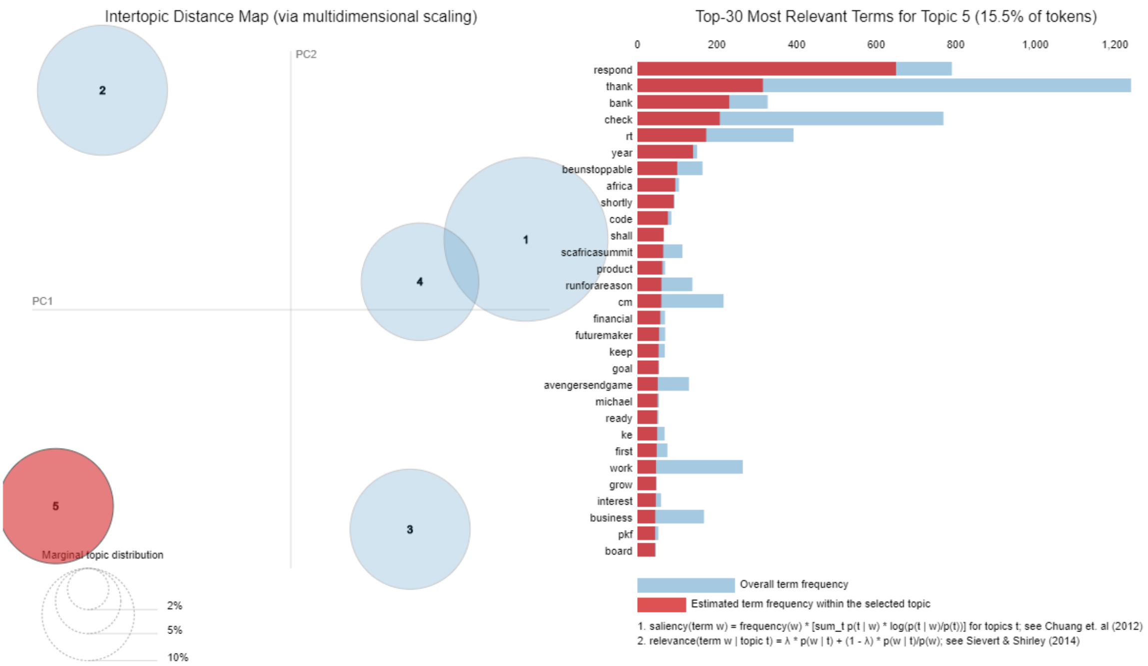
4.2.3 Topic 3



4.2.4: Topic 4



4.2.5 Topic 5



4.3 Topic frequency distribution

In LDA, as modelled by the Dirichlet distribution, a document may have multiple topics. Words, when considered outside a document, can also belong to multiple topics. Document 1 can vahe90% Topic 2 and 10% Topic while Document 2 can have 20% Topic 1 and 80%Topic 2 (Calvo-González, Eizmendi, & Reyes, 2018). Figure 3 shows the topic distribution based on the number of tweets. As shown the dataset has more distribution of Topic 3 with more than 3500 tweets while Topic 4 is the least less than 1500 tweets.

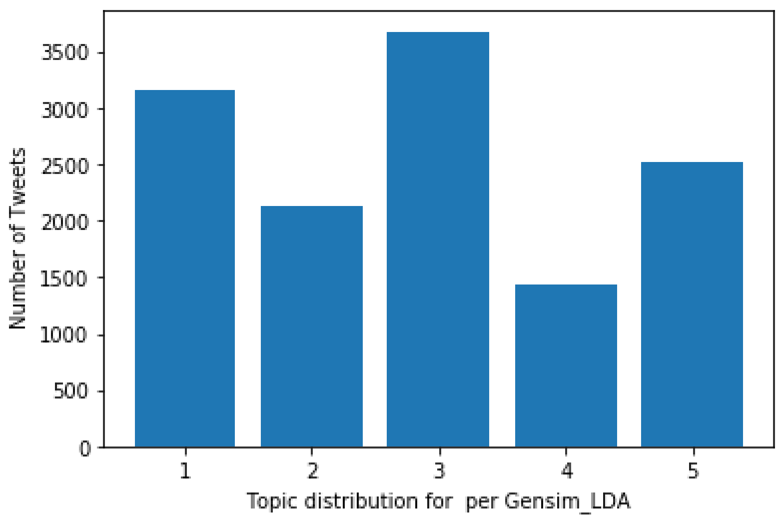


Figure 3: Gensim Topic Distribution

With social media, data are generated and disseminated in the public domain hence Sentiment analysis can be used as a monitoring tool for social media content (Namugera, Wesonga, & Jehopio, 2019) . Sentiment analysis involves classifying opinions in text into different levels of polarities like "positive" or "negative" or "neutral" (Barnaghi, Ghaffari, & Breslin, 2016). Polarity is the quantification of the sentiment with a positive or negative value (Kharde & Sonawane, 2016). From this study, most tweets (8000) had polarity of above 0.5 (positive). This implies that most interactions by the customers and banks were positive while around 500 tweets had a polarity of -0.5 (negative) and below.

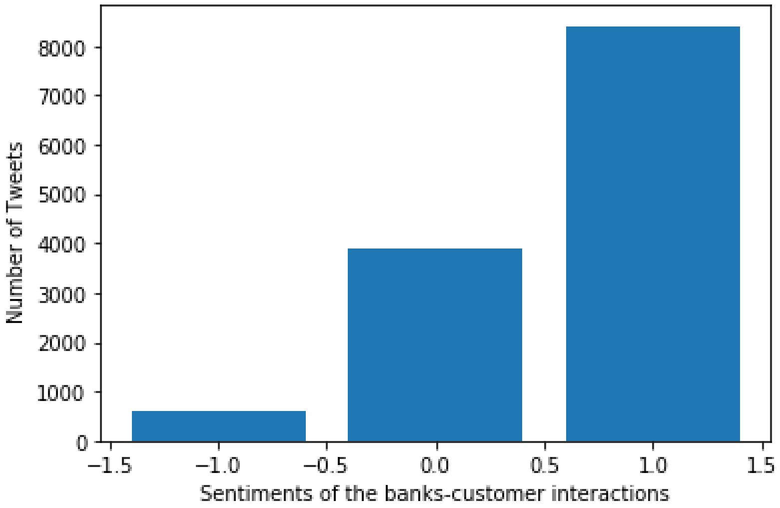


Figure 4: Polarity Distribution

5.0 Discussion

Social media interaction with customers makes it easier to market companies products / services while maintaining social connections with customers. Marketing through social networking sites has been found to be effective as current social media users are exposed to brand-related messages and are more likely to interact with a product. A successful process of profiling and segmentation requires a company to specify its business goals. Using approaches that represents targeting and measurement, management ought to agree and have clear goals at the start of any segmentation process. The practice of profiling and segmentation implies evaluating a various suitable approaches method suitable for the situation.

Customer retention is more critical than the acquisition of new customers. Twenty percent of customers contribute more to the company's revenue than the rest, according to the Pareto principle. A number of different consumer attributes can be used to identify customers. Customer segmentation can be performed using a variety of unique customer characteristic (Srivastava, 2016). One of the ways can be profiling and segmentation using online interactions as we have demonstrated. Through these banks and even other sectors, can tailor their marketing strategies can be tailored, plan product development plans, targeted advertising campaigns on various products and services (Christy, Umamakeswari, Priyatharsini, & Neyaa, 2018).

6.0 Conclusion

In this paper, we showed how organizations can infer latent insights from their interactions with customers. While most sentiments were positive, we were able to discover underlying themes behind these interactions. Most Companies have been able to manage to understand power of talking directly to their customers. This was shown by 'feedback' as one of the themes identified from analysis. Support responses on different queries were being responded in matter of minutes to customers by banking personnel. Loan applications processes and also debit and credit card issues that emanated from customers were being attended to by the banks. In cases where, a customer received delayed responses, they often complain and even urge other followers to boycott services of the bank. This shows that providing great customer service online gives a company is a platform to demonstrate their attention to their customers. If complaints or queries are also not addressed, it tends expose weakness of the company and at great lengths damage reputation chasing away potential customers. In addition, through this, competitors can be able to tap into disillusion customers from online interactions. In addition, the above analysis showed how social media data can be constructively used to devise targeted marketing strategies based on the dominant themes emanating from these online engagements.

Through topics and sentiments generated organizations can be able to profile customers and improve their user experiences in services they offer. This paper shows that that latent topic analysis can be an appropriate and effective method for analysing social media engagements and is a vital component for effective customer segmentation. Topic modelling and sentiment analysis does not only help build a customer profile but also supports customer relationship management and even can be used in recommender systems.

Acknowledgment
This study is one of the outcomes from the SIMElab Africa and was supported by NRF Grant NRF/1/MMC/418.

7.0 References

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. arXiv preprint arXiv:1707.02919.

An, J., Kwak, H., Jung, S. G., Salminen, J., & Jansen, B. J. (2018). Customer segmentation using online platforms: Isolating behavioral and demographic segments for persona creation via aggregated user data. Social Network Analysis and Mining, 8.

Balakrishnan, V., & Lloyd-Yemoh, E. (2014). Stemming and lemmatization: a comparison of retrieval performances.

Barnaghi, P., Ghaffari, P., & Breslin, J. G. (2016). Opinion mining and sentiment polarity on twitter and correlation between events and sentiment. In 2016 IEEE second international conference on big data computing service and applications (BigData). IEEE.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. The Annals of Applied Statistics, 1(1), 17-35.

Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent dirichlet allocation. Journal of Machine Learning, 3, 993-1022.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. Marketing Science, 35(6), 953-976.

Calvo-González, O., Eizmendi, A., & Reyes, G. (2018, Jan). Using Text Mining to Measure Policy Volatility. Policy Research Working Paper. World Bank Group.

Carrera-Trejo, J. V., Sidorov, G., Miranda-Jiménez, S., Moreno Ibarra, M., & Cadena Martínez, R. (2015). Latent Dirichlet Allocation complement in the vector space model for Multi-Label Text Classification. International Journal of Combinatorial Optimization Problems and Informatics, 6(1), 7-19.

Charoensukmongkol, P., & Sasatanun, P. (2017). Social media use for CRM and business performance satisfaction: The moderating roles of social skills and social media sales intensity. Asia Pacific Management Review, 22, 25-34.

Chen, Y. R. (2017). Perceived values of branded mobile media, consumer engagement, business-consumer relationship quality and purchase intention: A study of WeChat in China. Public Relations Review, 43(5), 945-954.

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2018). RFM ranking–An effective approach to customer segmentation. Journal of King Saud University-Computer and Information Sciences.

Coates, A., & Ng, A. Y. (2012). Learning feature representations with k-means. n Neural networks: Tricks of the trade, pp. 561-580.

Corrales, D. C., Ledezma, A., & Corrales, J. C. (2018). From theory to practice: A data quality framework for classification tasks. Symmetry, 10(7).

Deng, S., Lin, Y., Liu, Y., Chen, X., & Li, H. (2017). How Do Personality Traits Shape Information-Sharing Behaviour in Social Media? Exploring the Mediating Effect of Generalized Trust. Information Research: An International Electronic Journal,, 22(3).

Dolan, R., Conduit, J., Frethey-Bentham, C., Fahy, J., & Goodman, S. (2019). Social media engagement behavior. European Journal of Marketing.

Dolnicar, S. (2007). Market Segmentation in Tourism. In: Tourism. In: Tourism Management, Analysis, Behaviour and Strategy. Cambridge: CABI, pp. 151-173.

El Mehelmi, H., & Sadek, H. (2019). Investigating the usage of social customer relationship management (SCRM) and its impact on firm performance in the mobile telecommunication services: Egypt case. Journal of Business and Retail Management Research, 13(03), 282-291.

Elena, C. A. (2016). Social Media–A strategy in developing customer relationship management. Procedia Economics and Finance, 39, pp. 785-790.

Ezenkwu, C. P., Oloyede, A., Umana, I. S., Jerome, O., & Ekpo, E. (2013). Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services. In IEEE International Conference on Emerging & Sustainable Technologies for Power & ICT in a Developing Society (NIGERCON).

France, S. L., & Ghose, S. (2019). Marketing analytics: Methods, practice, implementation, and links to other fields. Expert Systems with Applications, 119, 456-475.

Haraty, R. A., Dimishkieh, M., & Masud, M. (2015). . An enhanced k-means clustering algorithm for pattern discovery in healthcare data. International Journal of distributed sensor networks, 11(6).

Hopkins , B. (2016, July 28). Insights-Driven Businesses Are Stealing Your Customers. Retrieved from Forrester: https://go.forrester.com/blogs/16-07-28-insights_driven_businesses_are_stealing_your_customers/

Hudson, S., Huang, L., Roth, M. S., & Madden, T. J. (2016). The influence of social media interactions on consumer–brand relationships: A three-country study of brand perceptions and marketing behaviors. International Journal of Research in Marketing, 33(1), 27-41.

Hughes , A. M. (1994). Strategic Database Marketing. Probus Publishing.

Im, Y., Park, J., Kim, M., & Park, K. (2019). Comparative Study on Perceived Trust of Topic Modeling Based on Affective Level of Educational Text. Applied sciences.

Islam, T. (2019). Yoga-Veganism: Correlation Mining of Twitter Health Data. arXiv preprint.

Jadczaková, V. (2013). Review of segmentation process in consumer markets. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 61(4), 1215-1224.

Kearney, M. W. (2016). rtweet-package: rtweet: Collecting Twitter data. Retrieved from Cran: https://rdrr.io/cran/rtweet/man/rtweet-package.html

Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. arXiv preprint arXiv:1601.06971.

Lanjewar, R., & Yadav, O. P. (2013). Understanding of Customer Profiling and Segmentation Using K-Means Clustering Method for Raipur Sahkari Dugdh Sangh Milk Products. International Journal of Research in Computer and Communication Technology,, 2(3), 103-107.

Lee, H., & Bae, I. (2016). A Study on Understanding Social Media users by Segmentation Analysis Focus on South Korea. Indian Journal of Science and Technology,, 9-43.

Litterio, A. M., Nantes, E. A., Larrosa, J. M., & Gómez, L. J. (2017). Marketing and social networks: a criterion for detecting opinion leaders. European Journal of Management and Business Economics.

Malik, U. (2019, April 09). Python for NLP: Topic Modeling. Retrieved from Stackabuse: https://stackabuse.com/python-for-nlp-topic-modeling/

Malthouse, E. C., Haenlein, M., Skiera , B., Wege, E., & Zhang, M. (2013). Managing customer relationships in the social media era: Introducing the social CRM house. Journal of interactive marketing, 27(4), 270-280.

Marisa, F., Ahmad, S. S., Yusof, Z. I., Hunaini, F., & Aziz, T. M. (2019). Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model. International Journal of Integrated Engineering, 11(3).

Mohammadian, M., & Makhani, I. (2016). RFM-Based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies. International academic journal of accounting and financial management,. International academic journal of accounting and financial management, 3(6), 21-35.

Mueller, G. (2019). Why Twitter is the Ideal Platform for Engagement. Retrieved from convince and convert: https://www.convinceandconvert.com/social-media-strategy/twitter-engagement/

Munar , A. M., & Jacobsen, K. S. (2014). Motivations for sharing tourism experiences through social media. Tourism Management, 46-54.

Namugera, F., Wesonga, R., & Jehopio, P. (2019). Text mining and determinants of sentiments: Twitter social media usage by traditional media houses in Uganda. Computational Social Networks, 6(1).

Paker, N., & Vural, C. A. (2016). Customer segmentation for marinas: Evaluating marinas as destinations. Tourism Management, 56, 156-171.

Paliouras, K., & Siakas, K. V. (2017). Social customer relationship management: a case study. International Journal of Entrepreneurial Knowledge, 5(1), 20-34.

Pansari, A., & Kumar, V. (2017). Customer engagement: the construct, antecedents, and consequences",. Journal of the Academy of Marketing Science, 45(3), 294-311.

Pater, A. M., Vári-Kakas, S., Poszet, O., & Pintea, I. G. (2019). P Segmenting Users of an Online Store Using Data Mining Techniques. In 2019 15th International Conference on Engineering of Modern Electric Systems (EMES) (pp. 205-208). IEEE. In 2019 15th International Conference on Engineering of Modern Electric Systems (EMES) (pp. 205-208). IEEE.

Peppers, D., & Rogers, M. (2011). Managing customer relationships: a strategic framework. Hoboken:. Hoboken:: John Wiley & Sons. Retrieved from http://proquestcombo.safaribooksonline.com/book/sales-andmarketing/ 9780470930182.

Pereira, J. F. (2017). Social media text processing and semantic analysis for smart cities. arXiv preprint arXiv:1709.03406.

Ransbotham, S., & Kiron, D. (2018). Using analytics to improve customer engagement. MIT Sloan Management Review.

Scridon, M. A. (2008). Understanding customers-profiling and segmentation. Management & Marketing-Craiova, 1, 175-184.

Sharp, D. E. (2003). Customer relationship management systems handbook. Auerbach publications.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces, (pp. 63-70).

Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. O'Reilly Media, Inc.

Sokolova, M., Huang, K., Matwin, S., Ramisch, J., Sazonova, V., Black, R., & Sambuli, N. (2016). Topic modelling and event identification from twitter textual data. arXiv preprint .

Srivastava, R. (2016). Identification of customer clusters using RFM model: a case of diverse purchaser classification. International Journal of Business Analytics and Intelligence.

Tajudeen, F. P., Jaafar, N. I., & Ainin, S. (2018). Understanding the impact of social media usage among organizations. Information & Management,, 55(3), 308-321.

Tajvidi, R., & Karami, A. (2017). The effect of social media on firm performance. Computers in Human Behavior.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. Journal of Marketing Research, 51(4), 463-679.

Trainor, K. J., Andzulis, J. M., Rapp, A., & Agnihotri, R. (2014). Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. Journal of Business Research, 1201-1208.

Voorveld, H. A., van Noort, G., Muntinga, D. G., & Bronner, F. (2018). Engagement with social media and social media advertising: The differentiating role of platform type. Journal of advertising, 47(1), 38-54.

Wind, Y. J., & Bell, D. R. (2008). Market segmentation. Routledge.

Wong, E., & Wei, Y. (2018). Customer online shopping experience data analytics: Integrated customer segmentation and customised services prediction model. International Journal of Retail & Distribution Management, 46(4), 406-420.

Wright, B., Nadler, S., Borders, A. L., Schwager,, P. H., & Sasnett, A. (2018). Corporate Social Media: A Typology of Consumers. Atlantic Marketing Journal, 7(1).

Wu, Y., & Wu, J. (2016). The impact of user review volume on consumers' willingness-to-pay: a consumer uncertainty perspective. Journal of Interactive Marketing, 33, 43-56.

Yang, C., Pan, S., Mahmud, J., Yang, H., & Srinivasan, P. (2015, September). Using personal traits for brand preference prediction. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing , pp. 86-96.

Zahrotun, L. (2017). Implementation of data mining technique for customer relationship management (CRM) on online shop tokodiapers.com with fuzzy c-means clustering. n: 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE). Yogyakarta.

Zhang, M., Guo, L., Hu, M., & Liu, W. (2017). Influence of customer engagement with company social networks on stickiness: Mediating effect of customer value creation. International Journal of Information Management,, 229-240.

Zhang, Y., & Pennacchiotti, M. (2013, May). Predicting purchase behaviors from social media. In Proceedings of the 22nd international conference on World Wide Web, pp. 1521-1532