

SNA, LDA AND SENTIMENT ANALYSIS FOR COMPETITIVE INTELLIGENCE IN AGRO –E COMMERCE

Afolabi, Ibukun T & Eniola Enilolobo Israel

Research Scholar, Department of Computer and Information sciences, Covenant University, Ota, Nigeria

ABSTRACT

Every e-commerce company yearns for increased sales. The decisions that could bring about this desire are based on several factors that are hard to determine by these companies until recently. The advent of e-commerce analytic techniques has made it easier to make these decisions by uncovering these factors from past data. Factors like customer sentiment and trending products, which influence company decisions, are now easy to evaluate using e-commerce analytic techniques. This research demonstrates several techniques useable by e-commerce companies to gain competitive intelligence. It shows how sentiment analysis, social network analysis, topic modelling can be used to gain valuable insights that will foster business growth and increase sales. This research made use of heterogeneous data from several social media platforms and Google trend data. It demonstrates these methods on a real-life case study and also suggests recommendations from the results. This research is valuable to e-commerce companies by providing and demonstrating methods and procedures required extracting adequate competitive intelligence inference from available data.

KEYWORDS: *E-Commerce Analytic Techniques, Adequate Competitive Intelligence Inference*

Article History

Received: 04 Mar 2022 | Revised: 29 Mar 2022 | Accepted: 30 Mar 2022

INTRODUCTION

E-commerce, fully known as electronic commerce, can be defined as purchasing and selling products and services over the internet. The history of e-commerce dates as far back as the late 1900s following the Electronic Data Interchange (EDI) invention in the 1960s, which was used in standardizing electronic transactions between customers and vendors, easing the process of data exchange between them. There are several other defining moments in the history of e-commerce, like Michael Aldrich's demonstration of the first online trading system in 1979. As time went on, the popularity of e-commerce only increased with more people participating in it daily. A study by Statista, a market and consumer data firm based in Germany, showed that about 1.66 billion people shopped online in 2017. The numbers have been increasing over the years, with 1.79 billion people shopping online in 2018, 1.92 billion people in 2019, and 2.05 billion forecasted to shop online in 2020 (Coppla, 2020). That's over a quarter of the world population predicted to shop online in just 2020. These numbers are not surprising given the numerous benefits e-commerce provides. These benefits include providing sellers with a platform to showcase their goods and services to a broad market, and consumers with the ease of shopping for a wide range of goods and services anytime and anywhere, even from the comfort of their homes.

Due to the massive growth of e-commerce, there was a need for way to understand the entire e-commerce market and identify the changes and trends in customers' behaviour, helping retailers make decisions that would promote their sales and bring about increased profit. This is where e-commerce analytics came in. E-commerce analytics is the process of collecting data from all aspects of your web shop and using the information to understand trends and changes in customer behaviour to make data-driven choices that will lead to more online purchases (Moore, 2020). There are about four types of e-commerce analytics, namely: Descriptive analytics, which answers the question “what exactly happened?”, Diagnostic analysis, which helps answer the question “why did that happen?”, Prescriptive analysis, which helps answer the question “what do we do when this happens?”, and Predictive analysis which helps answer the question “what is it that will likely happen?”.

E-commerce analytics can be applied to several areas. One central application area is for gaining competitive intelligence. Competitive intelligence, also known as corporate intelligence, is the capacity to obtain, evaluate, and apply information about rivals, consumers, and other market elements to help a company gain a competitive edge. In the case of e-commerce companies, this could be done via social media analysis since social media and the internet constitute the business environment of e-commerce firms. In this era, the effect social media has on the growth and development of brands is enormous. According to (Singh & Singh, 2018), marketing on social media impacts the building of brands and their sales. Also, the feedback gotten from customers via social media (like tweets about their products) further helps them design proper marketing strategies for their products. Social media engagement analysis may involve a couple of methods and techniques like sentiment analysis, which helps identify customers' inclination to products and services. Topic modelling and social network analysis can also be used to discover valuable insights for competitive advantage. These give the retailers enough information to provide their customers with the best products and services and gain a competitive edge over their competitors.

Marketing and promotional campaigns have changed themselves from being reliant on mass-market channels (like radios and television) to the social stage following the introduction of social media. This is particularly true for e-commerce companies, a lot of which boast of having numerous followers and a large fan base on their several social media platforms (Singh & Singh, 2018). All of that would be useless if they aren't able to gain insights and feedback on their products and services from those customers (or clients) who follow them. Some of these companies fail to completely satisfy their customers because they ignore the one platform from which they can get undiluted feedback on their products and services.

Social media has proven to be vital to the building of brands, especially in this age, and according to (Vinodhini & Chandrasekaran, 2017), online review research can assist retailers in forecasting revenue and has a driving impact on consumer purchasing decisions. This is why its effects need to be studied. It could help analyze the competitive environment and predict distribution and marketing plans even before the launch of products. Also, the feedback gotten on products already in the market from their customers could help in making improvements where necessary (Singh & Singh, 2018). This research is focused on demonstrating specific e-commerce analytic techniques for gaining competitive intelligence on the data gotten from the social media platforms of an agricultural ecommerce case study.

Competitive intelligence has proven over and over again to be very crucial in the success of any e-commerce business. From increased sales to increased business margins, to better customer retention, its benefits are numerous and undeniable (Avinash & Akarsha, 2017). Bearing these multiple benefits in mind, it can be concluded that competitive intelligence is the gateway to survival in the future because, it helps e-commerce firms analyze consumer behaviour and

predict their interests, helping these firms optimize their marketing strategy in real-time thus, creating value for themselves (Avinash & Akarsha, 2017).

LITERATURE REVIEW

This section is focused on reviewing past works that has to do with using social media for competitive advantage on the ecommerce platform. It identifies the methods and approaches used in each analysis, the distinct advantage each work has on improving business growth, their strengths, and validation methods. It also highlights the effectiveness and limitation of each method and points out the possible directions and factors to be considered if further research is to be conducted.

(Benoit & Van Den Poel, 2012) looked into the benefits of social network mining in terms of customer retention. It checks to see whether kinship network-based variables, in addition to traditional variables like socio-demographic and purchasing background, will increase the predictive ability of consumer retention models. The results of the evaluation showed that the extended churn model is more effective at distinguishing churners from non-churners (according to the AUC metrics) and that it's also better at forecasting customers at high-risk (according to the lift value) when compared with the traditional model. One major benefit of this study is that it presents the advantages of using network-based knowledge in churn management. It also reveals that database marketers should store network information in their data warehouses (if they haven't already) and use network-based data in their churn prediction models.(Benoit & Van Den Poel, 2012).

(Kaushik et al., 2018) investigated what effect e-WOM (Electronic Word of Mouth) or series of helpful reviews, as well as other review characteristics such as ratings, length, informativeness, and valence, have on the sales of products on the ecommerce platform.

The data used for this study was gathered from the Amazon. in platform on a variety of items since a rich collection of product reviews are made available on the Amazon website. A linear regression model was then run between the content of the product page and the following week's sales of the product. The findings of this study show that positive or helpful reviews and the order in which they appear on a website have a direct impact on a customer's decision to purchase a product and the sales of products. This information would aid content creators in coordinating review content on product sites. A major limitation of this study is that the data used for the analysis was based on just electronic products, so the focus was on just mobile products. Further research can work on comparing the impact of these review features across different product categories. (Kaushik et al., 2018).

(Chen & Lin, 2019) looked at the impact of social marketing campaigns and activities on user retention, purchase, and participation intentions using their social identity, perceived importance, and satisfaction. This study created a questionnaire that was initially completed by ten participants who had prior experience with social media marketing. The goal was to improve areas of the questionnaire that appeared to be too ambiguous or difficult to comprehend. This study then conducted a pilot test at a university in northern Taiwan where 46 valid sample responses were collected after this was completed and the first draft was improved. The Cronbach's alpha value for this pilot test was greater than 0.7, indicating that the questionnaire was stable and consistent internally. After obtaining these positive results, the study went ahead to invite social media users, via an online community, to fill the questionnaire on the chosen online questionnaire system. The sample size used for the data analysis was 502, 52 % of which were females and 48 % were males. The study then carried out Partial Least Squares (PLS) analysis using Smart PLS. The findings of this study showed that social media marketing activities have a strong impact on perceived value and social identification, which in turn has an impact on customer

satisfaction, retention, purchase intent, and participation intent. It also demonstrated to businesses that creating and managing an online brand community can boost business performance while preventing community members from defecting to competitors. A major limitation of this study is that it only considers the impact of social media marketing activities. Future research should look into all other aspects of modern social networking websites, as well as their effects on user demands and actions. (Chen & Lin, 2019).

Bhattacharyya & Bose in 2020 investigated the influence of Facebook likes on purchases and recommendations in a linked E-commerce site. Rather than relying on secondary data to investigate this relationship, this study used scenario-based controlled experiments to improve construct conceptualization precision and determine causality. Participants in the experiment were exposed to a Facebook advertisement with a 'shop now' button that would redirect them to the linked e-commerce website offering the product displayed in the advertisement. The number of Facebook likes and aggregated product ratings on the simulated e-commerce website were also being manipulated using two levels: low and high. This resulted in four treatment conditions because they tried to examine whether their proposed relationship would hold under low or high conditions of the product rating. The Latin-square method was then used to ensure that each participant received all of the treatments, with the treatment series being randomized among them. The results of this study showed that more Facebook likes positively affected a customer's decision to purchase or endorse a product. This study has a couple of limitations that could be improved on in further research. They include: the experiment used utilitarian and low-involvement items, and it was based on a single form of social cue (i.e. Facebook likes). Experimenting with a different class of products could help generate additional insights. Further studies on other social media-driven e-commerce purchases, such as Instagram-driven e-commerce, could be carried out to support this study's results. (Bhattacharyya & Bose, 2020).

(Jung & Jeong, 2020) used data science and machine learning models to assist start-up businesses with a social media presence in predicting the effect of their social media marketing efforts. Startups-list.com, a website that offers information on all start-ups and newly founded businesses was used for the study and twitter was the chosen social media platform, due to its popularity and prominence. VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis API for Python was used for the sentiment analysis of tweets because rather than simply presenting the ratings, it describes the extent to which a sentiment is positive or negative. These findings show that of all the models tested, deep learning is the best machine learning model for forecasting social media interaction levels. The results also show that the number of tweets sent by a start up company, followed by the number of likes and retweets on their posts, are the most significant factors in predicting the effectiveness of their social media marketing efforts. This shows that these small companies just need to publish and be visible enough on social media to see the impact of their efforts (Jung & Jeong, 2020).

(Zhan et al., 2020) proposed an analytical framework for retail pharmacy organizations to define the most-discussed issues by their customers on social media, identify areas that need enhancement based on observed negative responses, and evaluate the correlations between key principles to increase consumer satisfaction. This framework uses social media data to make recommendations about how to improve operations management and make better strategic decisions. The data for this analysis came from the Twitter pages of three of the main retail pharmacy organizations in the United Kingdom, including Superdrug, Boots, and Superdrug. To define the core topics being tweeted by online users, Latent Dirichlet Allocation (LDA) was used in the text mining method. This technique identified 56 topics that were being discussed by online consumers and by combining information from the retail pharmacy industry's literature analysis and human judgment, these topics were further divided into subjects. Some factors such as 'waiting time', 'product quality',

'delivery', 'marketing campaign', and 'product availability' that are specifically linked to retail pharmacy operations and service management were also identified. Amongst these factors, 'delivery' had the highest weight (i.e. 1.33), indicating that subjects on delivery were mostly discussed by the consumers and that it's a major issue in the retail pharmacy industry. The sentiment analysis was then performed with SentiStrength to explore the spectrum of feelings or sentiments (either positive, neutral, or negative) shared by different customers in each online post.

(Arora et al., 2020) performed competitive analysis on ecommerce site using promoted post detection on the social media platform. The research is an attempt to assist brands in identifying perspectives such as the disparity between marketed and organic posts. It provides a concise overview of a novel solution for the issue of promoted post identification. To conduct the study, the dataset was obtained hourly using graph API from the "brand pages" features offered by Facebook. The extracted data contained 10,685 brands from 84 industries. To train various slow learners, ensemble machine learning strategies such as bagging and boosting were used, and then the results were aggregated using majority voting and weighted average. In stacking, a logistic regression model in conjunction with a generalized linear model and ensemble models (i.e. Random forest and XG Boost) was used. The results of the study show that compelling content plays an important role in making a post a boosted one. The research is appropriate, precise, and important for all products and sectors that report a high acceptance rate of post-promotion identification models due to high results. Future works can involve applying deep learning principles to see if they can improve the system's performance and assign weight to the features with respect to their importance to label them as paid or organic. Other ensemble methods could also be applied to obtain better outcomes. (Arora et al., 2020).

(Holland et al., 2020) analyzed the relative output or success of individual firms among existing competitors using market-level data. They demonstrated how big data from consumer click streams can be used to create a new wave of B2B (Business to Business) analytical systems. The network structure showing the relationship between airlines and Online Travel Agents (OTAs) through a network of interconnected websites was first visualized using the collected online panel data. The methodology then extends the basic visualization by formally and mathematically identifying new theoretical constructs based on network diagrams derived from a market-level source/loss matrix. The research demonstrated how big data can be used to further understand emerging business models, with a focus on how consumer big data can be transformed and analyzed to aid B2B marketers in their decision-making. The study also highlights how businesses will be able to map out their business climate from the consumer's viewpoint, gaining new insights into the industry, especially regarding systemic knowledge, that web server software cannot provide. Improvement on this study by future works can be realized by integrating the distinctive characteristics of online panel data with other forms of business data, such as revenue data and competitive knowledge obtained by collaborating with managers from individual organizations within a network. (Holland et al., 2020).

Looking at how diverse the above-reviewed works and methods are, the importance of constant improvement in methodologies used for social media analytics in today's world has become very clear. The high rate at which social media activities are increasing, coupled with the value social media analytics adds to businesses demands the constant refinement of methodologies and approaches to analysis. A good number of the above-reviewed works support this, as the methods they employed were either modified or improved versions of previous ones. Today, with the constantly increasing need for social media analytics as a source of gaining business intelligence, it is equally important to constantly improve on methods that would obtain better and more accurate results than previous ones, since that would imply better business performance.

METHODOLOGY

Figure 1 show the research framework used in carrying out the competitive analysis. This framework includes several activities needed to be carried out, from acquiring the data to presenting recommendations after the data analysis has been completed. The first step required for our analysis is data discovery. It involves identifying the type of data that would be required for the analysis, the form in which it would be needed, and the source(s) from which it would be obtained. This step requires a proper understanding of the problem to be solved for it to be properly executed. For this research, the social media data of the selected case study and its competitors was retrieved from the Facebook and twitter. The data was mostly unstructured, so there would be a need for further processing after acquiring the data. The descriptive analytics data was retrieved from Google trends using the short listed the most searched for item on their website.

In the second stage which is data acquisition, data was gathered from different sources. There are several options for doing this, and they include: downloading the data, scraping the web, collecting data directly from a sponsor, or buying the data from aggregators or other sources. The case study used for this research project is Afrimash (www.afrimash.com), a Nigerian e-commerce company based in Ibadan, Oyo State, that specializes in delivering livestock products to farmers across the country. They are also one of the foremost online marketplaces for agricultural products. Data was extracted from their social media accounts on Facebook and Twitter, coupled with data extracted from the Facebook account of a few of their competitors. This data was used to carry out the analysis and provide actionable recommendations that could incite an increase in sales and eventually cause increased business growth.

Different data pre-processing techniques was carried out based on the type of data that was retrieved and the analytics methods to be used. Details of this are revealed in section 4.0. In the data visualization stage, the graphical representation of data was made using libraries in programming languages like Matplotlib in Python or other software like Tableau, Gephi, etc.

The data cleaning phase was concerned with the separation of irrelevant data from useful information. It is aimed at improving the accuracy and efficiency of analysis results by improving the quality of the data. It also aims to tackle some data-related issues such as data inconsistency, data incompleteness, noisy data, spam data, etc., prior to its use in the analysis. In the data transformation stage, the data was converted into the form in which it would be used during analysis. The tasks involved include privacy protection. This was necessary because users are posting a lot of sensitive knowledge on social media, whether they are mindful of it or not, such as their names, everyday life habits, financial and health statuses. The collection and release of social network data for research raises legitimate privacy issues(Wang, 2014). As a result, the obtained data needs to be carefully converted in order to remove sensitive information. For the network data, the data used was converted to adjacency matrix, edge lists, and adjacency lists.

In this research, model development was done using centrality metrics from social network analysis. In addition to this, the Latent Dirichlet Allocation (LDA) was used for topic modelling and sentiment analysis was done with sentiwordnet in RapidMiner studio (<https://rapidminer.com/>).

The use of social network visualization will help to graphically reflect identified networks in a clear and understandable way. Quantitative functions are heavily used in social network visualization tools to numerically describe different network attributes. These functions, also known as social network metrics, were built on the basis of everyday mathematics(Arif, 2015). Some of the social network analysis metrics to be used for the analysis of the social network are discussed below.

Centrality: Social network analysis focuses on 'relationships', and 'actors' are essential to all forms of relationships. As a result, actor attribute classification or profiling is an integral part of every social network study (Arif, 2015). Actors are network members who are either discrete entities (for example, health-care customers or neighbourhood residents) or group units (for example, health organizations within a community) linked by relational ties (Hawe et al., 2004). Centrality identifies the most influential actors or those who are heavily engaged in relationships with other network members. In layman's words, centrality refers to the importance of participants or actors in a network. The degree centrality of a given node in a network is the number of links or edges incident on it. It is used to classify nodes or actors with the most connections. For a graph $G = (V, E)$, the degree centrality for a particular node is given as:

$$C_D(v) = \text{deg}(v) \quad (1)$$

Where $C_D(v)$ is the degree centrality of the vertex or node (v), and $\text{deg}(v)$ is the number of edges incident on the vertex (v).

The degree centrality for the entire graph can be expressed as:

$$C_D(G) = \frac{\sum_{i=1}^{|V|} [C_D(v_i) - C_D(v_j)]}{H} \quad (2)$$

Eigenvector centrality, which is a more advanced variant of degree centrality. It is dependent not only on the number of incident connections but also on their quality. This implies that a node's centrality importance increases as a result of its relations with high-status nodes (Arif, 2015). Let $A = (a_{v,u})$ be the adjacency matrix of a graph G with V vertices and E edges. The A can be defined as follows:

$$A_{v,u} = \begin{cases} a_{v,u} = 1, & \text{if vertex 'v' is linked to vertex 'u'} \\ a_{v,u} = 0, & \text{otherwise} \end{cases}$$

A vertex's eigenvector centrality can be calculated using:

$$C_E(v) = \frac{1}{\lambda} \sum_{u \in N(v)} x_u = \frac{1}{\lambda} \sum_{u \in G} a_{v,u} x_u \quad (3)$$

Where λ is a constant and $N(v)$ is the set of neighbours of the vertex v .

Closeness centrality, which represents the degree of proximity (direct or indirect) between any node and the rest of the nodes in the network (Arif, 2015). Given a graph G with 'n' nodes, the closeness centrality of a vertex 'v' can be represented as:

$$C_c(v) = \frac{n-1}{\sum_{k=1}^n d(u_i, v)} \quad (4)$$

Where $d(u_i, v)$ is the geodesic distance between u_i and v .

Betweenness centrality, which quantifies the number of times a node serves as a bridge along the shortest path between two other nodes by measuring the proportion of all shortest paths that travel through it. Nodes with a strong

betweenness centrality play a critical role in the network's information flow and cohesiveness and are considered central and invaluable to the network because of their role in the network's information flow(Arif, 2015). The betweenness centrality of a vertex 'v' can be defined as follows:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{5}$$

Where $\sigma_{st}(v)$ is the number of paths that travel through 'v' and σ_{st} is the cumulative number of shortest paths from node 's' to 't'.

The sentiment analysis was carried out using the lexicon-based approach, which relies on the sentiment lexicon, or a list of documented and precompiled sentiment expressions to analyze the text. It could be dictionary-based (which is based on the discovery of opinion seed terms, followed by a review of their synonyms and antonyms in a dictionary) or corpus-based (which begins with a seed list of opinion terms, then searches a vast corpus for other opinion words to aid in the discovery of opinion words with context-specific orientations using statistical or semantic methods)(Medhat et al., 2014).

The topic modelling was done using a generative probabilistic machine learning model. It believes that each text is a mixture of a small number of topics and that each word's appearance is traceable to one of the document's topics.LDA is a three-level hierarchical Bayesian model that models each item in a collection or document as a finite mixture over an underlying set of topics. Each topic is modelled as an endless combination of topic probability. This is quite useful in the context of topic modelling, as these topic probabilities give an explicit representation of the document (Blei et al., 2003).

In the final stage of recommendation and presentation, the results of the analysis are graphically presented in the form of a dashboard using charts, graphs, etc. It would aid in the addition of a visual component, making the results much more convenient and simple to comprehend. Coupled with the presented results, actionable recommendations would also be provided. These recommendations stem from the proper understanding and interpretation of the analysis results and should aid decision-makers in making decisions that would bring about an increase in return on investments.

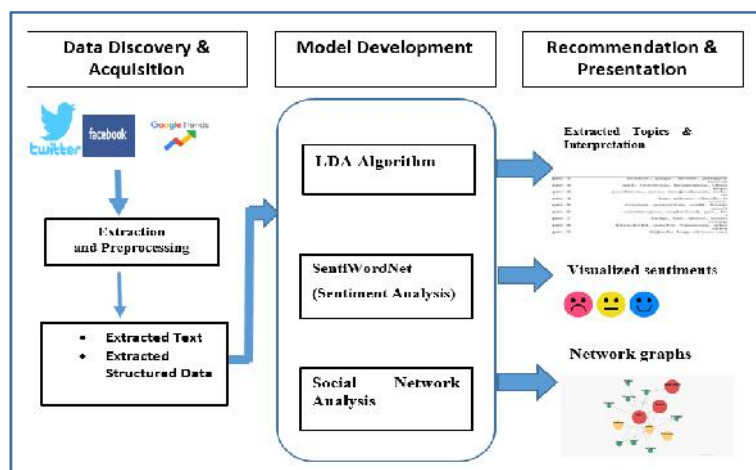


Figure 1: Methodology Framework.

ANALYSIS AND RESULTS

The analysis and result is presented based on the goals that are being addressed.

Goal 1: To obtain New Customers from the Competitive Environment

The first goal of our analysis was to study the competitive environment and obtain insights that could draw in more customers. This was carried out in two phases:

- Analyzing customer sentiments on the Facebook posts of the case study (Afrimash) and their competitors also.
- Performing topic modelling, also from the Facebook posts of the case study and their competitors.

In analyzing the sentiments of customers of both Afrimash and their competitors, data from their respective Facebook pages was gathered. Using Face Pager, a data-gathering tool, comments from their respective posts on Facebook within the last two years (from 22/02/2019 to 23/06/2021) were extracted. Also, Afrimash's website was scraped for comments and reviews on the products being sold there using Python and the Beautiful soup library. The data extracted from Face Pager was unfit for analysis as it was a direct copy of what the Face Pager interface looked like (that is, apart from the extracted data, it also comprised other unnecessary information like the object ID, key, type, query Id, etc.). To clean up the extracted data, Python was used. Using the pandas library, each of the extracted files was iterated over, identifying which of the entries were posts and which ones were comments. All the null values were dropped, and then the data frame was converted to a CSV file which was then saved. This pre processing process returned 1769 posts and 1241 comments for Afrimash, 152 posts and 845 comments for Day done, 1 post and 1 comment for Easy agro, 953 posts and 4974 comments for Farm crowdly, and 551 posts and 859 comments for Farm square. The comments extracted from each of the Facebook pages of the organization sand Afrimash's website were then imported into Rapid Miner and analyzed to extract the sentiment of their customers. In analyzing the customer sentiments, the text processing and word net extensions were first installed from the Rapid Miner marketplace. The actual sentiment analysis process then began by first tokenizing the text using the tokenize operator, then transforming the tokens to lower case using the 'transform case' operator, filtering the tokens by length (4 token 100) using 'filter tokens (by length)' operator, filtering the tokens by stop words using the 'filter stop words (English)' operator, generating a tri-gram using the 'generate n-grams (terms)' operator, stemming the tokens using the 'stem(wordnet)' operator, and finally, extracting the sentiment using the 'extract sentiment (English)' operator.

Figures 2, 3, 4, and 5 show the results of visualizing the extracted Facebook data. They show the distribution of the sentiments of Afrimash, Day done, Farm crowdly, and Farm square, respectively, using a line graph. The x-axis represents the number of comments from which the sentiments were extracted, while the y-axis represents the levels of sentiment (below zero is negative and above zero is positive).

More information on the sentiments is provided in Figure 6. It shows the aggregated sentiments of each of the organizations (Afrimash, Daydone, Farmcrowdy, and Farm square) for the period of time for which data was collected. This aggregation gives us an idea of the trend of sentiment over that period.

Figure 7 shows the distribution of sentiments in the extracted reviews from Afrimash's website using a bar chart.

The last activity carried out in the sentiment analysis process was extracting a couple of words from the negative sentiments in Afrimash's dataset of Facebook comments. Figure 8 shows some of the words that occur in the negative comments.

For the topic modelling, data from the extracted Facebook posts were used, and Latent Dirichlet Allocation, an algorithm for topic modelling, was implemented using Python on this data. The aim of this was to identify the topics most discussed by the respective organizations (Afrimash and its competitors). This was done using necessary Python modules and libraries required for the analysis. They include pandas, NLTK (for natural language processing), genism (for the topic modelling), re (for defining regular expressions), etc. The hyper parameter for the model (in this case, the number of topics) was set to three, and the following results as shown in figures 10, 11, 12 and 13 were obtained:

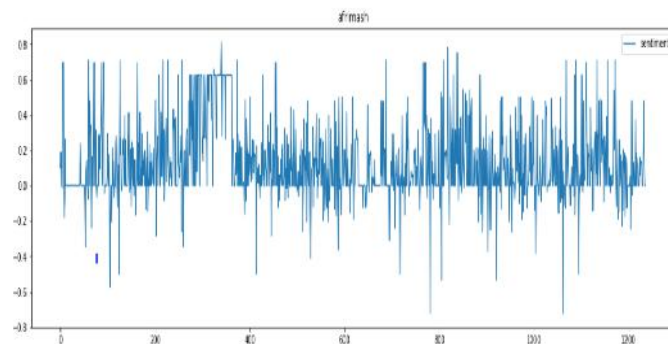


Figure 2: Line Graph Showing Sentiment Distribution of Afrimash's Facebook Comments.

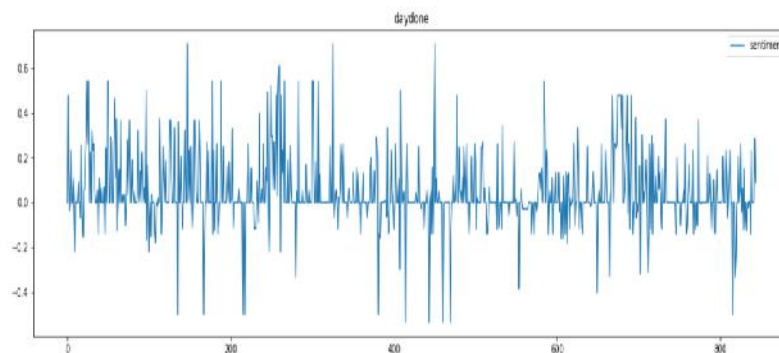


Figure 3: Line Graph Showing Sentiment Distribution of Daydone's Facebook Comments.

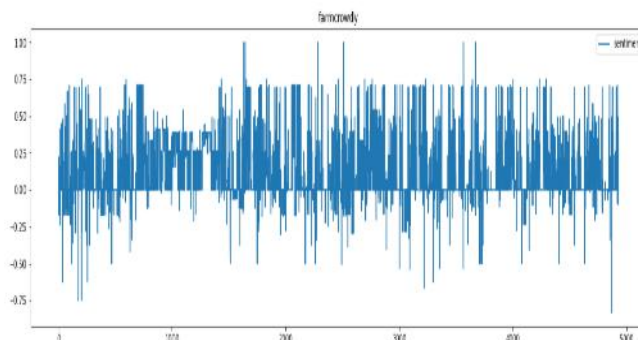


Figure 4: Line Graph Showing Sentiment Distribution of Farmcrowdy's Facebook Comments.

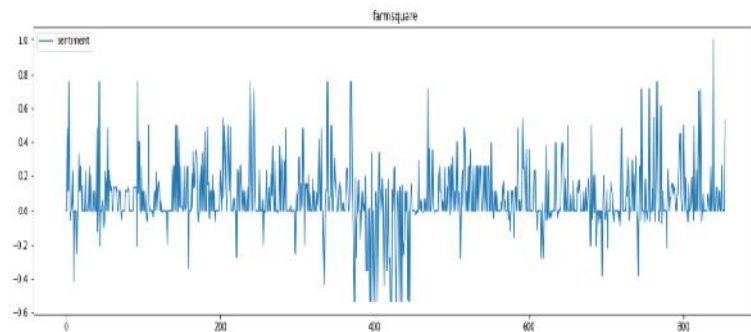


Figure 5: Line Graph Showing Sentiment Distribution of Farmsquare’s Facebook Comments.

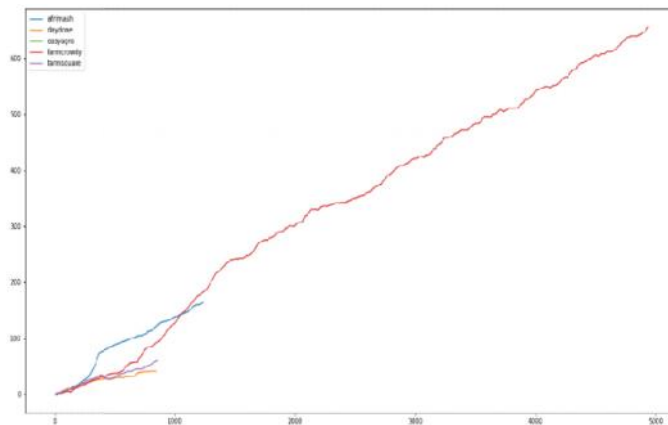


Figure 6: Aggregated Sentiments of Each of the E-Commerce Organizations.

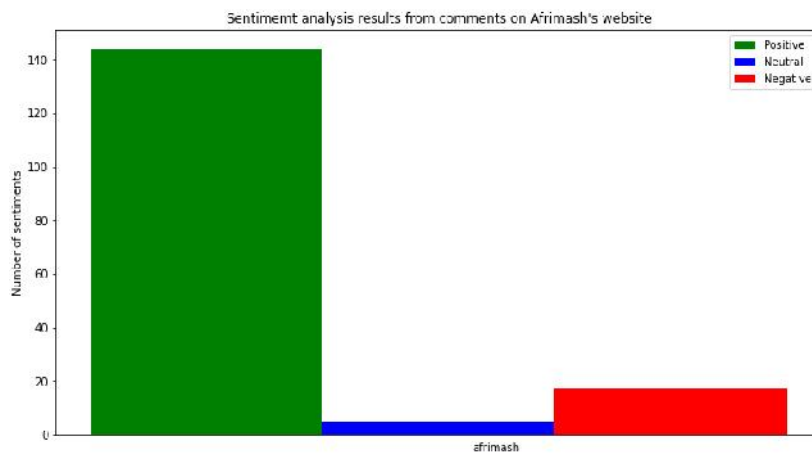


Figure 7: Bar Chart Showing Sentiment Distribution of Afrimash’s Website Reviews.

```
[ order ] / [ choose ] / [ inconvenience ] / [ using ] / [ link ] / [ king ] / [ chicken ] / [ cause ] / [ apology ] / [ turkey ] / [ sorry ] / [ sent ] / [ hello ] / [ buyer ] / [ audio ] / [ poor ] / [ chick ] / [ thank ] / [ accessible ] / [ place ] / [ site ] /
```

Figure 8: Some Negative Words That Appear in Afrimash’s Comments.

```
----- TOPICS IDENTIFIED: -----
----> (0, '0.023*afirimash' + 0.017*farming' + 0.016*training' + 0.014*register' + 0.011*online' + 0.010*visit' + 0.010*account' + 0.009*poultry' + 0.006*click' + 0.006*farmer'')
----> (1, '0.014*order' + 0.003*place' + 0.007*nationwide' + 0.007*chick' + 0.007*offer' + 0.006*machine' + 0.006*bird' + 0.006*broiler' + 0.006*click' + 0.006*crate'')
----> (2, '0.021*order' + 0.020*place' + 0.017*nationwide' + 0.015*deliver' + 0.009*poultry' + 0.007*bird' + 0.007*offer' + 0.006*water' + 0.005*product' + 0.005*capacity'')
```

Figure 9: Topics Identified from Afrimash’s Posts with Word Distribution.

```
----- TOPICS IDENTIFIED: -----
----> (0, '0.021*delivery' + 0.019*order' + 0.013*ibadan' + 0.012*available' + 0.012*doorstep' + 0.011*chicken' + 0.010*weekend' + 0.009*another' + 0.006*daydone' + 0.008*delivered'')
----> (1, '0.021*gerri' + 0.006*produce' + 0.006*thing' + 0.006*water' + 0.006*service' + 0.005*healthy' + 0.005*ecommerce' + 0.005*agric' + 0.005*product' + 0.005*think'')
----> (2, '0.013*agric' + 0.011*ginger' + 0.009*dream' + 0.009*produce' + 0.009*farmer' + 0.009*market' + 0.009*product' + 0.008*nigeria' + 0.007*ecommerce' + 0.006*farming'')
```

Figure 10: Topics Identified from Daydone’s Posts with Word Distribution.

```
----- TOPICS IDENTIFIED: -----
----> (0, '0.016*farmcrowdy' + 0.011*business' + 0.010*onyeka' + 0.009*amazing' + 0.008*nigeria' + 0.008*happy' + 0.008*africa' + 0.003*today' + 0.007*akumah' + 0.007*founder'')
----> (1, '0.016*farmcrowdy' + 0.014*agriculture' + 0.012*farmer' + 0.016*happy' + 0.007*nigeria' + 0.007*north' + 0.007*foods' + 0.007*monday' + 0.007*today' + 0.006*ginger'')
----> (2, '0.028*comment' + 0.023*section' + 0.013*answer' + 0.011*trivia' + 0.010*nigeria' + 0.009*farmer' + 0.009*world' + 0.008*amazing' + 0.007*fruit' + 0.006*thank'')
```

Figure 11: Topics Identified from Farmcrowdy’s Posts with Word Distribution.

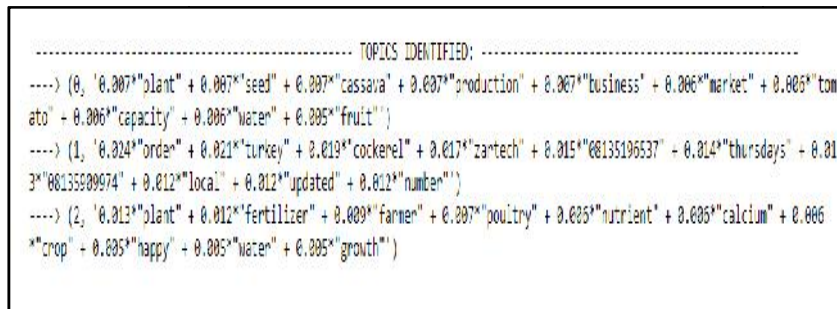


Figure 12: Topics Identified from Farm Square’s Posts with Word Distribution.

Goal 2: To Identify Best-Selling Products Using Descriptive Analysis

The data used for this simple descriptive analysis was obtained from Google trends by searching for certain keywords as they relate to the case study. These keywords include: antibiotic, battery cage, broiler feed, chi farms, drinker, eggs, feed formula, feed mill, feed production, fertilizer, fish feed, herbicide, layer feed, oil, point of lay, poultry, poultry equipment, red oil, scale, seedling, turkey, vet medicine, day-old chicks, incubators, and point of lay. The keywords used are the most frequent keyword searched for on the Afrimash website. A total of 25 keywords were used, and the aim of this analysis was to have an overview of how people have searched for these keywords within the past 12 months. The result of this analysis, shown in figure 14 gives an overview of the weekly search frequencies of the used keywords over the last twelve months. The variations in weekly searches can also be seen from this plot.

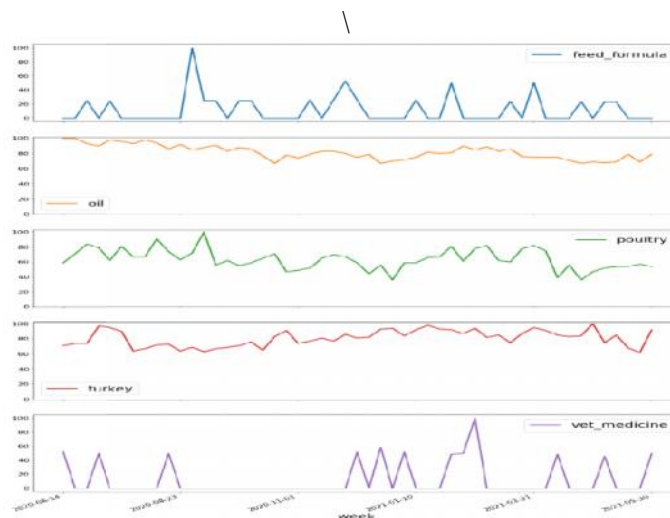


Figure 13: Line Graph Depicting the Number of Keyword Searches with Respect to Time (Only 5 Keywords is Displayed).

Goal 3: To Maximize Social Engagements using Social Media

The data used for this analysis was obtained from Twitter using Netlytics. The extraction was done using certain keywords and keyword combinations which include: Poultry; poultry, Nigeria; chicken, turkeys, layers, Nigeria. A total of three keyword combinations were used. Hence, three extracts were made. In the first extract using the keyword ‘Poultry’, the data contained users and who they replied to. In the second extract using the keyword combination ‘Poultry, Nigeria’, users’ replies to other users didn’t generate enough data for the analysis. Therefore, retweets, quotes, and mentions were also included in the search criteria.

In the last extract using the keyword combination ‘Chick, Turkey, Layers, Nigeria’, the data gathered from just user replies was also insufficient. Therefore, retweets, quotes, and mentions were also included in the search criteria used. The results of the network analysis visualized using different centrality measures is shown figures 15, 16, and 17. To further support the visualizations, the betweenness and closeness centrality of each of the networks were calculated.

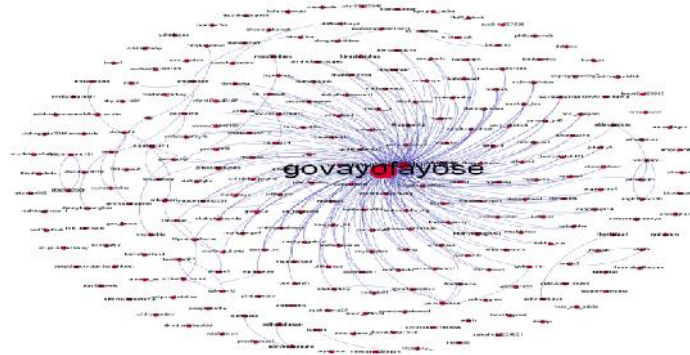


Figure 14: Network Generated from Gephi File Extracted Using ‘Chicks, Turkey, Layers, Nigeria’ As A Keyword.

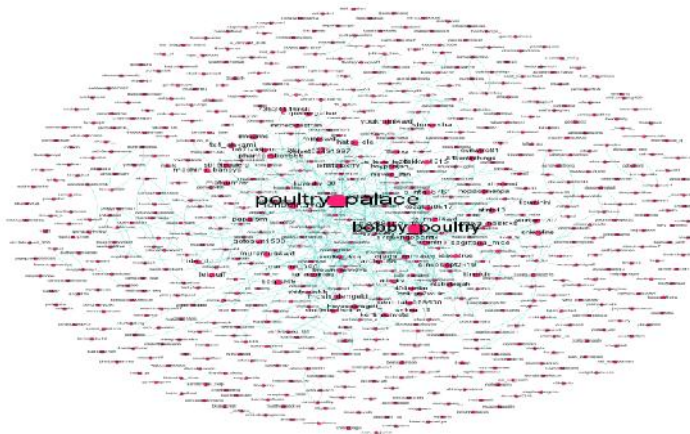


Figure 15: Network Generated from Gephi File Extracted Using ‘Poultry’ As A Keyword.

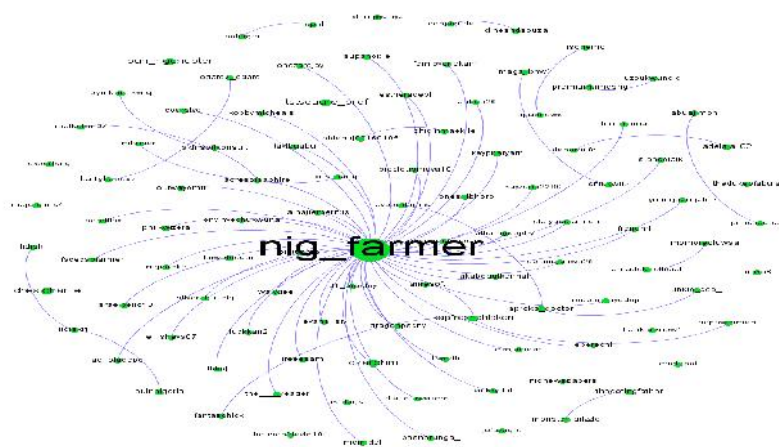


Figure 16: Network Generated from Gephi File Extracted Using ‘Poultry Nigeria’ As A Keyword.

DISCUSSIONS AND RECOMMENDATIONS

For the first goal which has to do with obtaining new customers from the competitive environment, the results of the sentiment analysis are pretty impressive, as figures 3, 4, 5 and 6 indicate that they have had more positive sentiments than negative. Figure 7, in particular, shows the growth of the positive sentiments over time, and although this is quite encouraging, a couple of negative sentiments were still identified. Figure 9 shows some of the words that occurred in the comments with negative sentiments. Words like ‘order’, ‘inconvenience’, ‘buyer’, etc. indicate that some customers are probably dissatisfied with the process of ordering products. Other words like ‘chicken’ and ‘turkey’ also probably indicate dissatisfaction with these particular poultry products. Afrimash might want to look into and improve on these areas in order to provide their customers with more satisfaction. The results of the topic modelling, as shown in figures 10, 11, 12 and 13 are a list of three topics being discussed on each of the Facebook pages of the organizations whose data was used for this project. Prior to the start of the main analysis, it was clear that of all the organizations, Farmcrowdy generated the most data. In comparison to Afrimash, they had fewer posts within the time frame the data was gathered but had way more engagement by their customers in the comment section. From the topic modelling results, we can tell from the third topic in the list that apart from business, Farmcrowdy engaged its customers in playing some games. This most likely kept them interested and active on Farmcrowdy’s page. Afrimash could also follow this initiative in order to have more engagements on their own Facebook page.

For the second goal which has to do with identifying best-selling products using descriptive analytics The result of the descriptive analysis, as shown in figure 14, shows that the following products, on average, have the most searches among customers: oil, feed formula, turkey, poultry, scale, etc. Afrimash could expand the production and availability of these products to leverage the needs of their customers, satisfying them and ultimately profiting from them. While production of products like feed formula, vet medicine, incubators, feed mill, etc., with low frequency of searches, could be strategically reduced to avoid losses. For the final goal which is on maximizing social engagements using social media, figure 15, 16 and 17 show the results of sizing the nodes (or users) based on their eigenvector centrality. As the eigenvector measure gives an idea of which nodes are important and valuable in the network, nodes like govayofayose, poultry_palace, bobby_poultry, and nig_farmer, which are shown in figures 15, 16, and 17 to have the highest eigenvector centralities, would be very crucial in marketing. These nodes could be contacted and reached out for effective marketing campaigns on Twitter as it relates to the sale of agricultural products (specifically those products used as keywords in the search criteria for obtaining the network data). Since the betweenness centrality measure indicates how much a node acts as a bridge to other nodes, it was discovered that nodes like govayofayose, uncerf, imosocials, bobby_poultry, poultry_0838, acresofsaphire, etc., with high betweenness centralities, would be very vital in reaching out to others in the network. Afrimash could take advantage of these nodes in getting across to other users in the network. The closeness centrality measure, on the other hand, measures how close a node is to other nodes. Nodes like cico39037424, smartobinna9, harphampeg, lidiski1, etc., high closeness measures could be leveraged on or collaborated with by Afrimash whenever the need for direct marketing arises, as they are closest to other nodes in the network.

CONCLUSION AND FURTHER WORK

In this research, we have been able to successfully demonstrate several techniques usable by any e-commerce organization in extracting useful insights about themselves. It combines the power of descriptive analysis, competitive analysis, sentiment analysis, topic modelling, and social network analysis to provide any e-commerce company with the ability to at any time understand how well their business is faring in the business environment. The framework demonstrated here is

easy to use as the tools, methods, and techniques used here are user-friendly. The results of the analysis are also easy to understand and make decisions from. They provide enough information with which certain business improvement decisions could be made. Even though the case study used in this research is an agro ecommerce business, it can easily be adapted to other ecommerce business.

The conclusions and findings of this study have provided e-commerce organizations with a method of obtaining actual business intelligence in the form of insights into the behaviour and wants of their customers. This information would enable them to draw out proper marketing plans, identify and target the correct clients, and deliver better services to drive customer satisfaction. They would be able to differentiate themselves from their competition and enjoy improved profit, company performance, and business growth as a result of this. The results and findings of this research would provide e-commerce companies in general, or more specifically, the case study used with tangible business intelligence in the form of insights into the behaviour and desires of their customers. This information would enable them to draw out proper marketing strategies, identify and target the right customers, and offer better services to drive customer satisfaction. With this, they would stand out from their competitors and enjoy increased profit, business performance, and business growth. In addition to this, they would also be able to identify any issues or weaknesses (if there are any) in the business model and also provide recommendations on how and where to make improvements. To improve upon this research, in the analysis done in goal 2, some of the keywords were seen to peak and trough at different times. Further works could study the relationship between the number of searches and its relationship with the period of search in order to gain a deeper insight into the search pattern of customers. Also, more keywords or keyword combinations could be used in gathering Twitter data for analysis in goal 3. This could provide richer data for the social network analysis, thus improving the chances of getting better results.

REFERENCES

1. Arora, A., Srivastava, A., & Bansal, S. (2020). *Journal of Retailing and Consumer Services Business competitive analysis using promoted post detection on social media. Journal of Retailing and Consumer Services, 54*(September 2019), 101941. <https://doi.org/10.1016/j.jretconser.2019.101941>
2. Avinash, B. M., & Akarsha, B. M. (2017). *Big Data Analytics for E-Commerce – Its Impact on Value Creation. International Journal of Advanced Research in Computer and Communication Engineering, 6*(12), 181–188. <https://doi.org/10.17148/IJARCCCE.2017.61235>
3. Benoit, D. F., & Van Den Poel, D. (2012). *Improving customer retention in financial services using kinship network information. Expert Systems with Applications, 39*(13), 11435–11442. <https://doi.org/10.1016/j.eswa.2012.04.016>
4. Bhattacharyya, S., & Bose, I. (2020). *S-commerce: Influence of Facebook likes on purchases and recommendations on a linked e-commerce site. Decision Support Systems, 138*(January), 113383. <https://doi.org/10.1016/j.dss.2020.113383>
5. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent Dirichlet allocation. Journal of Machine Learning Research, 3*(4–5), 993–1022. <https://doi.org/10.1016/b978-0-12-411519-4.00006-9>
6. Chen, S. C., & Lin, C. P. (2019). *Understanding the effect of social media marketing activities: The mediation of social identification, perceived value, and satisfaction. Technological Forecasting and Social Change, 140*(December 2018), 22–32. <https://doi.org/10.1016/j.techfore.2018.11.025>

7. Coppla, D. (2020). Number of digital buyers worldwide from 2014 to 2021. <https://www.statista.com/statistics/251666/number-of-digital-buyers-worldwide/>
8. Holland, C. P., Thornton, S. C., & Naudé, P. (2020). B2B analytics in the airline market: Harnessing the power of consumer big data. *Industrial Marketing Management*, 86(December 2019), 52–64. <https://doi.org/10.1016/j.indmarman.2019.11.002>
9. Jung, S. H., & Jeong, Y. J. (2020). Twitter data analytical methodology development for prediction of start-up firms' social media marketing level. *Technology in Society*, 63(November 2019), 101409. <https://doi.org/10.1016/j.techsoc.2020.101409>
10. Kaushik, K., Mishra, R., Rana, N. P., & Dwivedi, Y. K. (2018). Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon.in. *Journal of Retailing and Consumer Services*, 45(August), 21–32. <https://doi.org/10.1016/j.jretconser.2018.08.002>
11. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
12. Moore, K. (2020). E-commerce analytics 101: A marketer's guide to driving more online sales with data. <https://supermetrics.com/blog/e-commerce-analytics#:~:text=What is e-commerce analytics%3F, will drive more online sales.>
13. Singh, M., & Singh, G. (2018). Impact of social media on e-commerce. *International Journal of Engineering and Technology (UAE)*, 7(2), 21–26. <https://doi.org/10.14419/ijet.v7i2.30.13457>
14. Vinodhini, G., & Chandrasekaran, R. M. (2017). A sampling based sentiment mining approach for e-commerce applications. *Information Processing and Management*, 53(1), 223–236. <https://doi.org/10.1016/j.ipm.2016.08.003>
15. Wang, Y. (2014). *Data Preparation for Social Network Mining and Analysis [SINGAPORE MANAGEMENT UNIVERSITY]*. https://ink.library.smu.edu.sg/etd_coll/101
16. Zhan, Y., Han, R., Tse, M., Ali, M. H., & Hu, J. (2020). A social media analytic framework for improving operations and service management: A study of the retail pharmacy industry. *Technological Forecasting and Social Change*, 163(April), 120504. <https://doi.org/10.1016/j.techfore.2020.120504>

