International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Social Media Mining Methods Used to Improve Business Intelligence at Safaricom Public Limited Company

Caroline Kendi[1], Dr Collins Oduor[2]

[1]Master of Science in Applied Information Technology, Africa Nazarene University, Nairobi, Kenya

[2]Assistant Professor, Information Systems School of Science and Technology, United States International University , Nairobi, Kenya

Abstract: ***The study sought to establish social media mining techniques that were used at Safaricom PLC. This study was anchored on the tenets of the General Systems Theory and the Technology Acceptance Model (TAM) and took on a descriptive survey design. The target population in the study comprised of 150 employees of Safaricom PLC in the social media and business development departments. It used simple random sampling, which saw 72employees of Safaricom PLC being sampled. A questionnaire was used as the main research instrument. The reliability of the study was tested using the Cronbach Alpha test and the validity test was ensured by consulting experts and supervisors of the study. The collected data was sorted, cleaned and coded into SPSS 23 for subsequent descriptive and inferential statistics. The analyzed data was presented using charts, figures and tables. The study found that text mining was used for social media mining at Safaricom PLC as well as clustering and visualization. The study concluded that social media mining methods improved business intelligence and significantly predicted business intelligence at Safaricom PLC. The study recommends that the organization can share the information mined to other departments which can be used to show where it is not performing well and therefore improve its performance.***

Keywords: Business intelligence, social media mining, Safaricom Public Limited Company, social media mining techniques, social media

1. Introduction

1.1 Background of the Study

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information, information that can be used to increase revenue, cuts costs or both (Pippal *et al.*, 2014). It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically data mining is the process of finding correlations or patterns among dozens of fields in large relational databases (Rahman, 2012).

According to Kim and Leskovec (2011), social media has vast amount of user-generated data which can be utilized for data mining. Data mining of social media can amplify use of social media and perk up commercial intelligence to transport enhanced services. For example, data mining techniques can identify user sentiments for anticipatory preparation to develop suggestion systems for business of specific products and even to build new friendships or connect certain interest groups. Currently, Facebook uses likes, groups as well as posts of users to recommend users specific ads as well as new pages and groups.

Many organizations, individuals and even government of countries follow the activities on SM in order to obtain knowledge on how their audience reacts to postings that concerns them. Kairam, Wang and Leskovec (2012) noted that SM permits the effective collection of large scale data and this gives rise to major computational challenges. Nevertheless, the efficient mining of the information retrieved from these large-scale data helps to discover valuable knowledge of paramount importance in different fields like marketing, banking, government and defence. Again effectively mined SM data can be used as decision support tool by different entities that make use of SM contents for different purposes (Petrocelli, 2013).

Despite these recognized benefits and the wide range of vendor offerings and technology capabilities in the social media analytics space, businesses are still struggling with adopting, implementing and institutionalizing methodologies and techniques for an effective social media analytics program (Owyang, 2011). This study addresses this issue by advancing the viewpoint that social media analytics need to be positioned as a business intelligence practice – tying its various monitoring, discovery and predictive capabilities to the tactical execution of social media initiatives and to high level strategic objectives of the organization.

Currently businesses in Kenya are spending millions of money on social media advertising and social media marketing yet the efficacy of these approaches and their impact on the bottom line, customer base and brand awareness of these companies remains a mystery. This is the gap that this study intended to fill by determining social media mining techniques that would improve business intelligence and thus translate into measurable and impactful growth and development of businesses.

1.2 Purpose of the study

The general objective of this study was to establish the social media mining methods used to improve business intelligence at Safaricom Public Limited Company.

2. Theoretical Framework

This study was premised on the General Systems Theory propounded by Ludwig Von Bertalanffy in 1968 and the Technology Acceptance Model (TAM). According to the General Systems Theory, systems consist of more than just the sum of its parts and are made up of three parts, which are elements, interconnections, and purpose (Meadows, 2008). The main elements in general systems are stock, flow, and feedback. A stock is the history and record of information exchanged dependent upon flow and feedback (Meadows, 2008). The flow and feedback elements are known as the interconnections. When these elements and interconnections take place in an isolated environment it is called a closed system (Mulej, 2007). The last important part of systems theory is the purpose, which is known as the element behaviour or goal (Meadows, 2008).

Within the tenets of systems theory, social media mining provides stock information on the problems customers are having, what they enjoy about your product and what they would like from a product. The flow of feedback on these matters informs the purpose of the system which is to provide business intelligence to the companies that use social media mining. It is the interconnected nature of these elements that enables social media mining to provide actionable business intelligence to the management of the companies that use them. The feedback gathered from social media mining can be used to inform business intelligence which in turn inform strategic business decisions that can be used to increase revenue, lower costs, promote brand awareness and on board new clients.

The Technology Acceptance Model (TAM) was proposed by David Fred in 1986, this model helps in the explanation and prediction of the behaviour of the users of new technology; this model is an addition of the Theory of Reasoned Action (TRA) and explains how external variables such attitude, beliefs and intention of use influence the behaviour of users of technology (Dai & Kauffman, 2001). The theory posits that what determines usage of a new technology system is affected either directly or indirectly by the user's attitude, intentions and the user's perception of the usefulness of the system and its ease of use (Davila et al., 2003).

Over time, TAM has evolved and the original model has been extended into TAM2 to include aspects of social influence such as image, subjective norms and voluntariness into the explanation of perceived usefulness; cognitive instrumental processes such as result demonstrability, job relevance and output quality are also included in the TAM2 model (Davila et al., 2003). This new model has been tested in both mandatory and voluntary settings and the results strongly supported it since it led to 60% user adoption; this study will adopt TAM2 together with TAM as the baseline model (Davis & Venkantesh, 2000).

David and Venkantesh (2000) assert that the degree to which the person trusts that a system will advance their performance at work will determine if the individual will adopt the system or not, also the more the individual perceives the technology to be easy to use, the more accepted the technology will be by the users; conversely, if a technology is perceived to be complex or difficult to use, then its adoption rate will be slow.

This theory was suitable to this study since the usage of social media mining to improve business intelligence is a fairly new phenomenon worldwide, these strategies are mostly simple and easy to use by those who are technologically savvy (Petrocelli, 2013). However, TAM asserts that the adoption of a technology, in this case social media mining to improve business intelligence, is determined by the perception that the user has on its usefulness and ease of usage (Davis & Venkantesh, 2000). Therefore, the ease of usage of social media mining and the perception business people and business managers have on it determines the adoption of this innovation (Davis & Venkantesh, 2000).

2.1 Literature Review

Rahman (2012) contends that data mining techniques are capable of handling the three dominant disputes with SM data which are size, noise and dynamism. SM data sets are very voluminous and require automated information processing for analysing it within a reasonable time. As data mining also require huge data sets to mine remarkable patterns from data, SM sites appear to be perfect sites to work on especially where opinion/sentiment expression is involved (Petrocelli, 2013). SM data sets are also characterised by noisy data such as spam blogs and irrelevant tweets in case of twitters. The dynamism in SM data sets causes it to evolve rapidly over time and data mining techniques are versatile in handling such dynamic data (Owyang, 2011). Kaplan and Haenlim (2012) report that the most commonly used social media mining techniques are text mining, clustering and visualization.

2.2 Text Mining

A study carried out by Romero and Ventura (2010) established that text mining is an emerging technology that attempts to extract meaningful information from unstructured textual data. Text mining is an extension of data mining to textual data Text mining is focused on finding useful models, trends, patterns, or rules from unstructured textual data such as text files, HTML files, chat messages and emails (Chiang, Lin, & Chen, 2011). As an automated technique, text mining can be used to efficiently and systematically identify, extract, manage, integrate, and exploit knowledge from texts (Ananiadou, 2008). Different from traditional content analysis, text mining is mainly data driven and its main purpose is to automatically identify hidden patterns or trends in the data (Tsantis & Castellani, 2011) and then create interpretation or models that explain interesting patterns and trends in the textual data.

Many researchers have successfully used text mining techniques to analyze large amounts of textual data in business (Ingvaldsen & Gulla, 2012), health science (Li, Ge, Zhou, & Valerdi, 2012) and educational domains (Hung, 2012). Witten, Don, Dewsnip, and Tablan (2003) used text mining techniques to extract metadata from documents in a digital library and to enrich documents by marking up appropriate items in the text. They found that text mining can add additional values to the documents stored in the digital library and enrich the user experience. Tane, Schmitz, and Stumme (2004) used text mining to group e-learning resources and documents according to the similarities among different topics. Abdous and He (2011) used text mining techniques to analyze the online questions posted by video streaming students and identified a number of learning patterns and technology-related issues. Fuller, Biros, and Delen (2011) used text mining to detect deception and lies in real world data

2.3 Clustering

Jain (2010) reports that clustering is an automatic process that divides volumes of documents into groups that are related to each other on the basis of common themes or properties; it is a method to find out what a collection actually contains. An example of clustering is to analyse customers' emails and find the one's overlooked bearing a common theme. The goal of this analysis is to find a set of clusters having lesser intra-cluster similarity than inter-cluster similarity and finally try to maximize it (Gan, Ma & Wu, 2012). Thus, clustering eases the browsing process to find related information by identifying hidden similarities and over viewing contents of large documents.

Clustering is applied to identify properties of a set of data as date, cost etc. and divide them into clusters. These clusters or subsets can be used to find hidden similarities, provide brief on large database collection and simplify browsing process to find related or linking information (Ingvaldsen & Gulla, 2012). Clustering is used by Text Miner search engine which is supported by a robust algorithm to find groups that are more similar than other members of same or other groups. Text mining uses Hierarchical Clustering for textual data where it merges two similar clusters into one. This process continues until the final root cluster is obtained (Tsantis & Castellani, 2011). So, in this process of hierarchy both inter and intra-cluster properties are easily revealed.

2.4 Visualization

Marakas (2013) reports that data visualization is the process by which textual or numerical data are converted into meaningful images, data mining algorithms can figure out hidden data patterns as well. The reason why the data visualization can help on data mining is that the human brain is very effective in recognizing large amounts of graphical representations (Ware, 2010). Hence, if the visualization techniques can correctly convert the raw data into visual graphs, users can very likely detect the patterns hidden in text and numbers. As an alternative to mechanical data mining algorithms, visual exploration has proven as an effective tool in data mining and knowledge discovery (Wang et al., 2010). Data structural features can be effectively recognized by data seekers using data visualization (Nabney et al., 2015).

This process of recognizing patterns through human brain can facilitate users to understand the meaning of patterns more intuitively. Therefore, visualization can complement the data mining techniques (Wang et al., 2010). The combination of data mining and data visualization, plus the enormous storage space in data warehouse, can provide precious information to business decision makers today. Information and Scientific Visualization Data visualization is accepted as the new name of this discipline which consisted of two existing sub-areas: information visualization and scientific visualization (Post et al., 2003). The study of scientific visualization was officially launched through a research recommendation made by the National Science Foundation (NSF) in 1987 (Ma, 2011). Approximately the same time, the emerging data warehouse and data mining (Han & Kamber, 2010) paved the way for information visualization to apply on high dimensional business datasets. In general, the variables in a typical scientific visualization task are continuous and are about volumes, surfaces, etc. Information visualization tasks are apropos of categorical variables and the recognition of patterns, clusters, trends, outliers, and gaps (Shneiderman, 2013). A typical data mining task in a business data warehouse context is more related to information visualization.

3. Methodology

The study adopted a descriptive research design. Kotzab, Seuring, Muller and Reiner (2005) state that a descriptive design determines those involved in the research topic, what their role is, when, where and how these subjects are affected by the research topic and the implications if this study is not carried out. The Safaricom PLC was a desirable study site since it had a strong social media presence across Facebook and Twitter and it was in constant communication with its clients so as to assure quality and solve their problems. Nyambu (2013) reported that Safaricom is the leading telecommunication company that has adopted social media mining. The target population in the study comprised of 150 employees of Safaricom PLC in the social media and business development departments.

The study used simple random sampling, which saw 72employees of Safaricom PLC being sampled. A questionnaire was used as the main research instrument. The reliability of the study was tested using the Cronbach Alpha test and the validity test was ensured by consulting experts and supervisors of the study. The study realized an overall Cronbach Alpha coefficient of 0.845, which meant that the research instruments were reliable and could be used to collect data from the field.The collected data was sorted, cleaned and coded into SPSS 23 for subsequent descriptive and inferential statistics. Descriptive statistics were conducted through frequency counts, percentages, means and standard deviations to capture the distribution of responses on the key issues addressed in the study objectives. Inferential statistics were conducted using correlation and regression analysis. The analyzed data was presented using charts, figures and tables.

4. Findings

4.1 Demographic Information

This section presents the demographic characteristics of the respondents. The demographic information sought from the respondents in this study included: Response rate, gender, department of respondents, their highest level of education, as well as the age of the respondents.

The study sought to collect data from 72 employees from the business development and social media departments at Safaricom PLC. However, the study did not achieve 100% response as there were non-response incidents encountered during data collection. Therefore, out of the targeted 72 respondents, 64 returned the questionnaires successfully, which accounts for 89% of the response. The study hence achieved a response rate of 89% and a non-response rate of 11% as shown in Figure 1. This response was excellent as per Mugenda and Mugenda (2003) since it postulates a response rate that is above 70% and is sufficient for analysis.

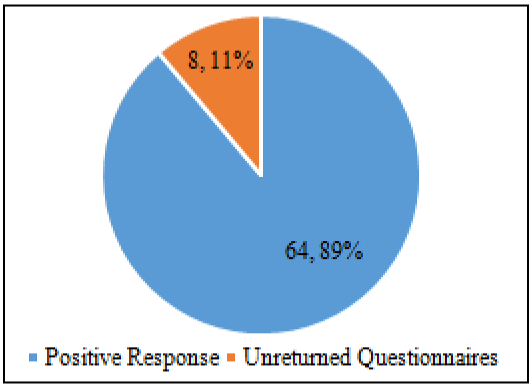

Figure 1: Response Rate
Source: Author Data, 2018

The findings on the gender of respondents indicated that 62% of the respondents were male while 38% were female. The findings indicate that majority of the employees at Safaricom PLC social media and business developments are male. The findings are shown in Figure 2.

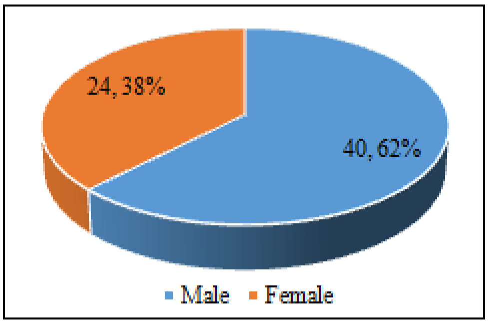

Figure 2: Gender of Respondents
Source: Author Data, 2018

The study found that 70% of the responses were obtained from employees in the social media department and 30% from the business development department. The findings imply more employees in social media department than business development department at Safaricom PLC a shown in Figure 3.

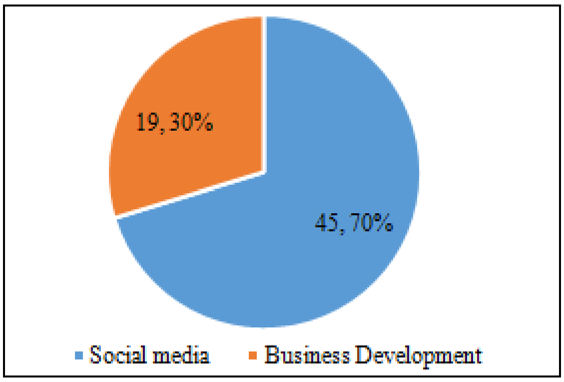

Figure 3: Department of Respondents
Source: Author Data, 2018

The study found that 66% of the respondents had attained Bachelor's degree as the highest level of education while 34% had attained a Masters' degree implying majority of the employees in the social media and business development departments were Bachelor's degree holders. Figure 4 presents these findings.

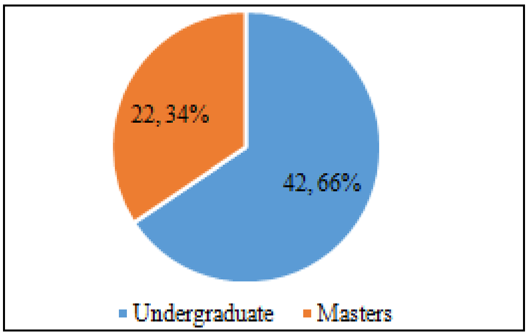

Figure 4: Highest Level of Education Attained
Source: Author Data, 2018

The study finally sought to determine the age of the respondents who participated in the study. The findings obtained indicated that 69% were aged 26-35 years, 22% were aged 36-45 years and 9% were aged 20-25 years. The findings imply that majority of the employees in social media and business development departments are between 25 and 35 years of age. The findings are shown in Figure 5.

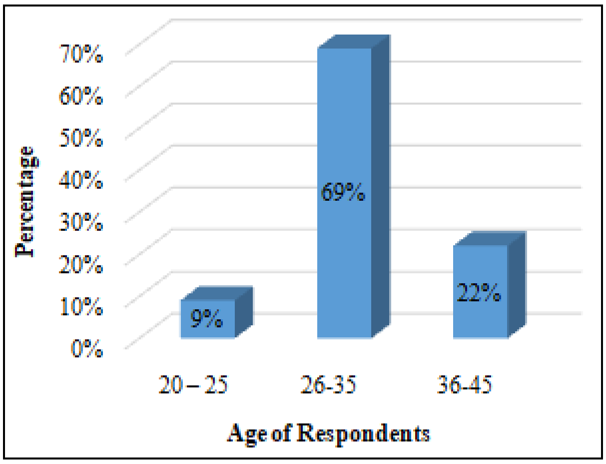

Figure 5: Age Range of Respondents
Source: Author Data, 2018

4.2 Presentation of the Findings

The aim of the study was to establish the social media mining methods used to improve business intelligence at Safaricom Public Limited Company. The findings are presented in Table 1.

Table 1: Social Media Mining Methods used at Safaricom PLC

		Strongly disagree	Disagree	Moderate	Agree	Strongly agree
Text mining is used for social media mining at my company	F	0	4	12	41	7
	%	0.0%	6.2%	18.8%	64.1%	10.9%
Clustering is used for social media mining at my company	F	0	7	13	30	14
	%	0.0%	10.9%	20.3%	46.9%	21.9%
Visualization is used for social media mining at my company	F	1	6	11	32	14
	%	1.6%	9.4%	17.2%	50.0%	21.9%

Source: Author Data, 2018

The findings of the study indicated that 64.1% of the respondents agreed that text mining is used for social media mining at Safaricom PLC, 18.8% neither agreed nor disagreed, 10.9% strongly agreed while 6.2% disagreed. Further, 46.9% of the respondents agreed that clustering is used for social media mining at Safaricom PLC, 21.9% strongly agreed, 20.3% neither agreed nor disagreed while 10.9% disagreed. In addition, 50% of the respondents agreed that visualization is used for social media mining at Safaricom PLC, 21.9% strongly agreed, 17.2% neither agreed nor disagreed, 9.4% disagreed while 1.6% strongly disagreed.

Means and standard deviations were used in the study to show the average response of the study and their deviations from each other respectively. The study also sought to analyse the views of the respondents on social media mining methods using a table of means and standard deviations. A Likert scale data was collected rating the views in a scale of 1 to 5 where 1 represented strongly disagree, 2 represented disagree, 3 represented neither agree nor disagree, 4 represented agree whereas 5 represented strongly agree. The results from the collected responses were analysed based on means and their standard deviations to show the variability of the individual responses from the overall mean. The findings are shown in Table 2.

Table 2: Means and Standard Deviations on Social Media Mining Methods

	N	Mean	Std. Deviation
Text mining is used for social media mining at my company	64	3.80	.717
Clustering is used for social media mining at my company	64	3.80	.912
Visualization is used for social media mining at my company	64	3.81	.941

Source: Author Data, 2018

The findings of the study indicate that the majority of the respondents agreed that text mining is used for social media mining at Safaricom PLC (M = 3.80, SD = 0.717); clustering is used for social media mining at Safaricom PLC (M = 3.80, SD = 0.912); and that visualization is used for social media mining at Safaricom PLC (M = 3.81, SD = 0.941). The standard deviations were less than 1 implying that the responses given by the respondents had a low variation from the mean value. This implies that the respondents had similar opinions regarding social media mining methods at Safaricom PLC.

The study also correlated social media mining methods with the various aspects of Business Intelligence (BI) at Safaricom PLC. BI was measured using increase in revenue, customer relationships and brand awareness at Safaricom PLC. Social media mining methods were measured using text mining, clustering and visualization. The correlation test was conducted at 5% significance level (2-tailed) and therefore the significance level was set at 0.025 above which the association is deemed insignificant and vice versa. Table 3 shows these findings.

Table 3: Correlation between Social Media Mining Methods and Business Intelligence

		Text Mining	Clustering	Visualization
Revenue increase	Pearson Correlation	.285**	.533**	.325**
	Sig. (2-tailed)	.006	.003	.004
Customer relationship	Pearson Correlation	.594**	.158*	.538**
	Sig. (2-tailed)	.000	.013	.005
Brand awareness	Pearson Correlation	.370*	.511**	.560**
	Sig. (2-tailed)	.021	.005	.003
*. Correlation is significant at the 0.05 level (2-tailed).				
**. Correlation is significant at the 0.01 level (2-tailed).				

Source: Author Data, 2018

The study determined that text mining significantly influenced customer relationship (r = .594, p<0.025), more than revenue increase (r = .285, p<0.025) and brand awareness (r = .370, p<0.025). In addition, it was found that clustering affected revenue increase (r = .533, p<0.025) more than customer relationship (r = .158, p<0.025) and brand awareness (r = .511, p<0.025). Finally, visualization affected brand awareness (r = .560, p<0.025) more than revenue increase (r = .325, p<0.025) and customer relationship (r = .538, p<0.025). The findings of the study imply that text mining is an important component of ensuring customer relationship, so was clustering in revenue realization and visualization in creating brand awareness.

Hypothesis Testing

The first null hypothesis of the study was stated as follows;
H_O: Social media mining methods do not improve business intelligence at Safaricom Public Limited Company.

Table 4: Model Summary for Social Media Mining Methods

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.617[a]	.447	.432	.130
a. Predictors: (Constant), Social Media Mining Methods				

Source: Author Data, 2018

The study found that social media mining methods explained 44.7% of the proportion in business intelligence at Safaricom PLC, R^2= .447. The results are presented in Table 4. The findings imply that other factors not studied in the current study account for 55.3% of the proportion in business intelligence in Safaricom PLC.

Table 5: ANOVA for Social Media Mining Methods

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.210	1	1.210	3.052	.006[b]
	Residual	24.590	62	.397		
	Total	25.800	63			
a. Dependent Variable: Business Intelligence						
b. Predictors: (Constant), Social Media Mining Methods						

Source: Author Data, 2018

The ANOVA findings indicate the reliability of the model on the relationship between social media mining methods and business intelligence at Safaricom PLC. The study found a significant value of 0.006 which is less than 0.05 at 95% confidence level and F value of 3.052. The regression model was therefore reliable. The findings are presented in Table 5.

Table 6: Regression Coefficients for Social Media Mining Methods

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.651	.364		8.246	.000
	Social Media Mining Methods	.622	.147	.617	3.747	.006
a. Dependent Variable: Business Intelligence						

Source: Author Data, 2018

The study found that social media mining methods significantly predicted business intelligence, β = .617, t = 3.747, p = .006. The p value was less than 0.05 and t value more than 1.96. These findings implied rejection of the null hypothesis. Therefore, the study concluded that social media mining methods improved business intelligence at Safaricom PLC. The results are presented in Table 6.

Adopting a linear regression model: $\mathbf{Y = \alpha + \beta 1X1 + \varepsilon}$
Where;
Where Y = Business Intelligence
β_0= Intercept
β_1 = Slope coefficients representing the influence of social media mining methods
X_1 = Social Media Mining Methods
ε = Error term

The linear equation for the study will therefore be:
Y= 4.651 + 0.622X1 + 0.364

The findings imply that for every unit increase in social media mining methods, business intelligence increases by 0.622; implying a positive effect of social media mining methods on business intelligence.

5. Discussion

The study determined that text mining was one of the main methods used by Safaricom PLC in business intelligence as shown in Table 1 and Table 2. In line with these findings, Romero and Ventura (2010) agreed that text mining is an emerging technology that attempts to extract meaningful information from unstructured textual data in many organizations. Ananiadou (2008) supports the findings that text mining can be used to efficiently and systematically to identify, extract, manage, and exploit knowledge from texts in various organizations.

In addition, the study determined that text mining is data driven and its used to automatically identify hidden patterns or trends in the data as shown in the correlation Table 3. These findings align to those posited by Tsantis and Castellani (2011) who identified text mining to create interpretation or models that explain interesting patterns and trends in the textual data.

Clustering was another social media mining method identified in Safaricom PLC in the study as shown in Table 1 and Table 2. In line with the findings of the study, Jain (2010) reported that clustering is useful in easing the browsing process to find related information by identifying hidden similarities and over viewing contents of large documents. Ingvaldsen and Gulla (2012) further supported that clustering is used in organizations to identify properties of a set of data as date that can be used to simplify browsing process.

The other social media data mining method in Safaricom PLC was visualization as shown in Table 1 and Table 2. Marakas (2013) and Ware (2010) also found that the visualization helps organizations in recognizing large amounts of graphical representations. Hence, users detect the patterns hidden in text and numbers. Wang et al. (2010) also agreed that visual exploration has proven as an effective tool in data mining and knowledge discovery.

The study found that various aspects of social media mining affected business intelligence aspects of customer relationship, brand awareness and revenue increase as presented in Table 3. In line with these findings, Post et al. (2003) posited that the combination of data mining and data visualization, plus the enormous storage space in data warehouse, can provide precious information to business decision makers today. Shneiderman (2013) also posited that the variables in a typical scientific visualization task are continuous and are about volumes and surfaces. A typical data mining task in a business data warehouse context is more related to information visualization. This is in line with the findings of this study.

6. Conclusions and Recommendations

The study tested hypothesis and found that social media mining methods improved business intelligence at Safaricom PLC. The study concluded that social media mining methods significantly predicted business intelligence. The study also concluded that the social media mining methods used at Safaricom weretext mining, clustering and visualization.

The study recommends that the organization can share the information mined to other departments which can be used to show where it is not performing well and therefore improve its performance. Further, the information mined is kept confidential. The study recommends that the organization can share part of the information to the public and give their customers chance to comment on the areas they feel they need to improve, and thereby improving both performance and business intelligence.

References

[1] Abdous, M. H., & He, T. (2011). Exploring cloud computing for distance learning. *Online Journal of Distance Learning Administration*, *14*(3), 1-10.

[2] Ananiadou, S. (2008). *National centre for text mining: introduction to tools for researchers.* Retrieved February 8, 2009.

[3] Chiang, D. M. H., Lin, C. P., & Chen, M. C. (2011). The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres. *Enterprise Information Systems*, *5*(2), 219-234.

[4] Dai, Q., & Kauffman, R. J. (2001, January). Business models for Internet-based e-procurement systems and B2B electronic markets: an exploratory assessment. In *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on* (pp. 10-pp). IEEE.

[5] Davila, A., Gupta, M., & Palmer, R. (2003). Moving procurement systems to the internet: The adoption and use of e-procurement technology models. *European management journal*, *21*(1), 11-23.

[6] Davis, G. H., &Venkantesh, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, *13*(1), 210-230.

[7] Fuller, C. M., Biros, D. P., &Delen, D. (2011). An investigation of data and text mining methods for real world deception detection. *Expert Systems with Applications*, *38*(7), 8392-8398.

[8] Gan, G., Ma, C., & Wu, J. (2007). *Data clustering: theory, algorithms, and applications* (Vol. 20). Siam.

[9] Han, J., &Kamber, M. (2010). Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems). *Morgan Kaufmann*.

[10] Hung, J. L. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology*, *43*(1), 5-16.

[11] Ingvaldsen, J. E., &Gulla, J. A. (2012). Industrial application of semantic process mining. *Enterprise Information Systems*, *6*(2), 139-163.

[12] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651-666.

[13] Kairam, S. R., Wang, D. J., &Leskovec, J. (2012, February). The life and death of online groups: Predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 673-682). ACM.

[14] Kaplan, A. M., &Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), 59-68.

[15] Kim, M., &Leskovec, J. (2011). Modeling social networks with node attributes using the multiplicative attribute graph model. *arXiv preprint arXiv:1106.5053*.

[16] Kotzab, H., Seuring, S., Müller, M., & Reiner, G.(2005). *Is there a right research design for your supply chain study*. Research Methodologies in Supply Chain Management, 1-12.

[17] Li, L., Ge, R. L., Zhou, S. M., &Valerdi, R. (2012). Guest EditorialIntegrated Healthcare Information Systems. *IEEE Transactions on Information Technology in Biomedicine*, *16*(4), 515-517.

[18] Ma, K. L. (2011). Guest Editor's Introduction: Large-Scale Data Visualization. *IEEE Computer Graphics and Applications*, *21*(4), 22-23.

[19] Marakas, G. M. (2013). *Modern data warehousing, mining, and visualization: core concepts* (pp. 100-101). Upper Saddle River, NJ: Prentice Hall.

[20] Meadows, H. (2008). Business intelligence: an analysis of the literature. *Information Systems Management*, *25*(2), 121-131.

[21] Mugenda O. & Mugenda A. G, (2003). *Research methods; quantitative and qualitative approaches.* Nairobi: *Acts press.*

[22] Mulej, W. (2011). Text mining in social networks. In *Social network data analytics* (pp. 353-378). Springer, Boston, MA.

[23] Nabney, I. T., Sun, Y., Tino, P., &Kabán, A. (2015). Semisupervised learning of hierarchical latent trait models for data visualization. *IEEE transactions on knowledge and data engineering*, *17*(3), 384-400.

[24] Nyambu, E. M. (2013). Influence of Social Media Marketing on Performance of Telecommunication Firms in Kenya. *International Journal of Innovative Research and Development*, *2*(9), 11-21.

[25] Owyang, J. (2011). How corporations should prioritize social business budgets. *Altimeter Group, San Mateo*.

[26] Petrocelli, T. (2013). *Social media marketing & analytics: Market landscape report*. Enterprise Strategy Group.

[27] Pippal, S., Batra, L., Krishna, A., Gupta, H., & Arora, K. (2014). Data mining in social networking sites: A social media mining approach to generate effective business strategies. *International Journal of Innovations & Advancement in Computer Science*, *3*(1), 14-22.

[28] Post, F. H., & Nielson, G. M., & Bonneau, G. P. (Eds.). (2012). *Data visualization: The state of the art* (Vol. 713). Springer Science & Business Media.

[29] Rahman, M. M. (2012). Mining social data to extract intellectual knowledge. *arXiv preprint arXiv:1209.5345*.

[30] Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601-618.

[31] Shneiderman, B. (2013). Why not make interfaces better than 3D reality?. *IEEE Computer Graphics and Applications*, *23*(6), 12-15.

[32] Tane, J., Schmitz, C., &Stumme, G. (2004, May). Semantic resource management for the web: an e-

learning application. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters* (pp. 1-10). ACM.

[33] Tsantis, L., & Castellani, J. (2011). Enhancing learning environments through solution-based knowledge discovery tools: Forecasting for self-perpetuating systemic reform. *Journal of Special Education Technology*, *16*(4), 39-52.

[34] Wang, Y., Luo, L., Freedman, M. T., & Kung, S. Y. (2010). Probabilistic principal component subspaces: a hierarchical finite mixture model for data visualization. *IEEE Transactions on Neural Networks*, *11*(3), 625-636.

[35] Ware, C. (2010). *Visual thinking: For design*. Morgan Kaufmann.

[36] Witten, I. H., Don, K. J., Dewsnip, M., &Tablan, V. (2004). Text mining in a digital library. *International Journal on Digital Libraries*, *4*(1), 56-59.