SOCIAL MEDIA AS A SOURCE OF PREDICTIVE ANALYTICS FOR CUSTOMER SATISFACTION: AN EMPIRICAL INVESTIGATION OF STANDARD GAUGE RAILWAYS USERS IN KENYA

Moses Kizito Njagi Ntale, Dr. Fr. Paul Mathenge and Barnabas Gikonyo
The Catholic University Of Eastern Africa School Of Business

ABSTRACT
Predictive analytics is used to analyze the vast amounts of information generated through internal and external sources such as live public transit data, train schedules, and bus feeds. The collection, storage, and mining of big data is on an increase as more automated platforms come online and this is an issue that is gaining attention in all business environments raising privacy concerns. However, the amount of data and its variety in data analytics may cause data management issues in areas of data quality, consistency and governance; resulting from different platforms and data stores in big data architecture causes data silos. Furthermore, integrating big data tools into a cohesive architecture that meets an organization's big data analytics needs is a challenging proposition for the analytics experts, which have to identify the right mix of technologies and then put the pieces together. This study therefore, investigated the influence of social media as a source of predictive analytics on customer satisfaction of Standard Gauge Railways (SGR) users. This research followed a cross sectional survey research design. The study targeted the customers and employees of SGR Nairobi terminus from which a sample size of 68 respondents was picked using from the Nairobi Terminus station through use of convenient sampling technique. This study used a questionnaire to collect primary data. Data obtained from the field was converted into useful information using qualitative and quantitative description qualitative was done through observation and analyzed through use of content analysis. On the other hand, quantitative data was analyzed through inferential techniques namely correlation and multiple regression. The findings indicated that usage of social media for information, mode of payment, SGR classes, and rates/fare had significant effect on customer satisfaction. It was recommended that the management of SGR should: place regular offers of discounts or freebies and give away on its sites; frequently update its social media sites with interesting information and product updates; and make the social sites more interactive to allow online members to invite others who are non-members of these social sites to sign up for the SGR services.
.
Keyword: Big Data, Data Analytics, Predictive Analytics, Social Media, Customer Satisfaction

1. INTRODUCTION
Information Technology has brought about break-through in science and technology which has led to tremendous developments to our economy (Siegel, 2013). Organizational progress is

heavily reliant on the adoption of these new developments in all areas of organization management (Michèle, 2010). As a result of the advancement in technology that has been embraced at a swift pace, information technology performs a major part in the administration of different organizational functions especially in the human resource sector (Coker, 2014). In the past, timely information has not been comprehensively and efficiently available as they are today (Schönberger & Cukier, 2013; Watson, 2014).

Predictive analytics projects source data from various different sources, the most common source is within the organization itself; other sources may include data purchased from third parties including transaction data, such as customer purchases; customer profiles entered information during sign up; data on whether customers responded to advertisements; pattern of customer web clicks; customer online interactions such as chats, surveys, and customer service calls; telemetric sensors and smart meters; subscriptions and lastly Social media such as Twitter and Facebook which is the most readily available (Moreira-Matias, et. al., 2016). By compiling data from varying data sources for developing predictive models, an organization gains a better overall perspective of its customers hence being in a better position to satisfy them (Ramachandran & Malladi, 2013)

Predictive analytics is used to analyze the vast amounts of information generated through internal and external sources such as live public transit data, train schedules, and bus feeds (Capgemi, 2014). Such insights procured through the use of predictive analytics can be utilized to plan routes, reduce traffic and transport bottlenecks (Letouze, 2011). Russell (2011) acknowledges that making fundamental changes in rail transportation is extraordinarily challenging. The transport industry as a whole is conservative about adopting new processes and methods of operation given the magnitude of potential loses if something goes wrong. Kenya lags behind in the quality and in technology deficient in the field of railway infrastructure; the Kenya Railway has a huge maintenance backlog that has left huge stretches of rail unusable and serious inefficiencies in key positions. These existing drawbacks have to be addressed to ensure that the Madaraka express is efficient and maximizes on returns as an investment and to ensure growth in the development and use of rail transport. This requires connections to other towns, modernization to digital signaling systems, Electrification, improved container wagons for block trains and mixed trains, effective and efficient train operations and better management of cross border rail services. Transport economists and regulators are faced with specific problems from the railway industry that are not particularly shared by other transport modes. Since SGR transports a wide range of products that vary in size, weight and shape; the particular cost of structure of railway companies; the key role played by infrastructure and networks; the existence of limited divisibility in input and output and rail transport as a public service.

The collection, storage, and mining of big data is on an increase as more automated platforms come online and this is an issue that is gaining attention in all business environments raising privacy concerns: What data should the government and organizations be allowed to collect and what safeguards should be in place about how it is used (Marz, 2015). Finlay (2014) explains that regulatory requirements may force organizations to work around data privacy provisions and governance which can interfere with cross-border transfer of data. Big data analytics ability to
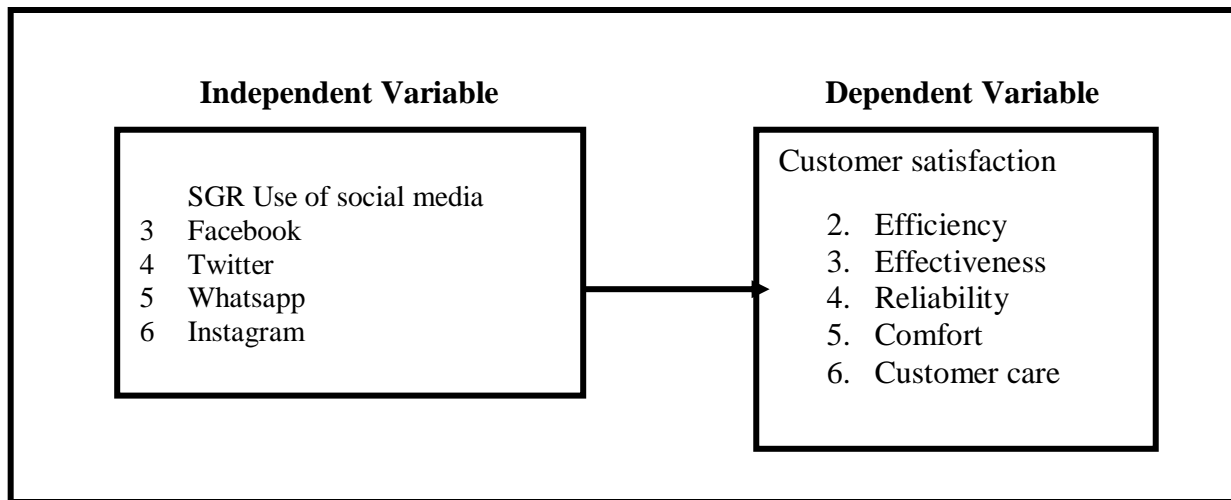
find connections in scattered data sets could potentially be used to deduct highly personal information (Korn, 2011). Dumbill (2012) observed that most people have very little understanding and concern about how organizations are using big data. However, as individuals understand the potential uses better, their concerns increase quickly. This suggests that as companies increasingly use big data analytics on customer data, the public is likely to become more concerned (Lindert, 2014).

The amount of data and its variety in data analytics may cause data management issues in areas of data quality, consistency and governance; resulting from different platforms and data stores in big data architecture causes data silos (Capgemi, 2016). Furthermore, integrating big data tools into a cohesive architecture that meets an organization's big data analytics needs is a challenging proposition for the analytics experts, which have to identify the right mix of technologies and then put the pieces together (Shaunak, et. al, 2017). This Study therefore intended to address these bottlenecks by establishing the relationship between predictive analytics and rail transport in Kenya in order to come up with viable recommendation.

2. CONCEPTUAL FRAMEWORK

From the empirical studies on BDA technology and its adoption in various industries and organizations we have presented a model based on the conceptual and determinants of adoption intentions as shown in Figure 1. In this study the conceptual framework is composed of determinants of big data analytics at the standard gauge railway transport of Kenya. Literature tells us that transport management authorities world over obtain insights from the real-time information or live data which help them to find the best alternatives to get from one destination to the other; thereby, reducing delays, and ensuring customer satisfaction. The conceptual framework illustrates that SGR management use predictive analytics solutions to enhance performance of their mode of transport.

Although predictive analytics tools determine the performance of SGR transport, this can be controlled by SGR attributes which are believed to contribute to SGR performance as illustrated. For this study, SGR attributes refer to the fundamental facilities and systems serving the railway transport, like trains schedules, rates, convenient destinations/stations, range of classes available. It is important to note that *SGR* travelers from Nairobi to Mombasa have to consider trains schedules, rates, convenience of the destinations/stations, range of classes available and speed, SGR attributes are offered in three categories including normal, economy, and business *class, p*lus the convenience of online booking.

Independent Variable	Dependent Variable
SGR Use of social media 3 Facebook 4 Twitter 5 Whatsapp 6 Instagram	Customer satisfaction 2. Efficiency 3. Effectiveness 4. Reliability 5. Comfort 6. Customer care

Figure 1: Conceptual Framework

Source: Author, 2019

3. MATERIALS AND METHODS

This research followed a cross sectional survey research design which was carried on management of SGR. The study targeted the customers and employees of SGR Nairobi terminus in Nairobi because they are the key stake holders of the SGR transport. A sample of respondents was drawn from the passengers waiting to board or those alighting from Nairobi Terminus station. The choice of the location for the study was drawn from the fact the Nairobi Terminus is one of the two major SGR terminus. The actual number of customers using the SGR services at any given time was not known therefore the following formula was used as suggested by Webster (1995) to estimate the sample size of this kind of population.

$$n = \frac{z^2 \pi (1 - \pi)}{(error)^2}$$

Where π is taken to be 50% of frequent users of SGR with Kenya Railways able to capture their data and the other 50% of unique users where Kenya Railways does not have any data on them. At the 90% desired level of confidence and margin error of 10%, the sample size (n) was calculated as follows:

$$n = \frac{(1.65)^2(0.5)^2}{(0.1)^2} = 68\ respondents$$

Convenient sampling technique was used to get the 68 respondents at the Nairobi Terminus station.

This study used primary data collected using a questionnaire as this was the best way to get the respondents opinions. Data obtained from the field was converted into useful information using qualitative and quantitative description. Data analysis was done using inferential and descriptive statistics. Descriptive statistics gave magnitude of the predictive analytics data sources presented in form of graphs and tables. Correlation analysis was used to estimate the association of the selected data sources and customer satisfaction. The following regression equation was used to estimate the influence of data sources for predictive analytics on customer satisfaction.

$CS = \beta_0 + \beta_1 DSPA + \beta_2 SGRA + \varepsilon$

Where CS is Customer Satisfaction at Madaraka Express

DSPA = Data Sources for Predictive Analytics

SGRA = SGR Attributes

ε = Error Term

β = Coefficients of the regression model

4. STUDY FINDINGS

Qualitative Description

It was established that SGR passengers were happy with the fact that the Madaraka Express had increased the frequency and improved travel experience on the populous route between Nairobi and Mombasa and is set to do the same on the western route to Maai-Mahiu terminus. Through interaction, it was noted that Passengers traveling on the *Madaraka Express* can book their tickets over the internet following the introduction of the online booking system. Madaraka Express introduced an SMS based payment platform utilizing USSD code *639# for passengers to book and pay for tickets through their mobile phones. The code enables SGR passengers to reserve seats, pay and even cancel tickets for the Madaraka Express.
Many passengers complained of the issue of ticket optimization which seems to not have been well thought out by the project management team, though passengers noted the issue was improving. Furthermore in practice major infrastructure projects are intense and have blind spots that may often take years to complete and understand. This means other elements critical to service delivery are still being streamlined and tested. Its interesting to note that while the SGR

management realizes that going digital does not mean having a mobile payment solution, the position Madaraka express finds itself offers a good opportunity to learn what it takes to compete in the information age while still relaying on a well thought out strategy.

Many SGR passengers acknowledged that Multimodal transit experiences are an everyday encounter in Kenya therefore, Madaraka Express structuring linkages to their terminals would be key to the service experience. It can be argued that designing a flawless corporate entity is impossible however; offering seamless connections for transit feels like a no brainer.

Quantitative Statistics

The model Summary of the regression model shows that 81.9% of the level of customer satisfaction variations can be explained by Rates/ Fare, social media, mode of payment, Schedules, stations and classes

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.934[a]	0.872	0.819	0.185474

a. Predictors: (Constant), Rates/fare, Social media platforms, Mode of payment, Schedules, Usage of social media for information, Destinations/stations, and Classes

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	3.975	7	0.568	16.508	0.000[b]
	Residual	0.585	17	0.034		
	Total	4.560	24			

a. Dependent Variable: Service satisfaction

b. Predictors: (Constant), Rates/fare, List of social media, Mode of payment, Schedules, Usage of social media for information, Destinations/stations, Classes

From the ANOVA table, the P-value (Sig) of the F statistics is 0.000, meaning that Predictive analytics data sources does have a significant effect on customer satisfaction of SGR passengers.

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.406	0.316		4.449	0.000	0.739	2.073
	List of social media	-0.085	0.063	-0.129	-1.358	0.192	-0.217	0.047

Usage of social media for information	0.702	0.149	0.805	4.713	0.000	0.388	1.016
Mode of payment	-0.253	0.052	-0.634	-4.893	0.000	-0.362	-0.144
Schedules	0.112	0.073	0.297	1.537	0.143	-0.042	0.265
Destinations/stations	0.022	0.061	0.070	0.369	0.716	-0.106	0.151
Classes	0.172	0.082	0.556	2.090	0.052	-0.002	0.345
Rates/fare	-0.328	0.084	-0.919	-3.894	0.001	-0.505	-0.150

a. Dependent Variable: Service satisfaction

From the coefficients of the regression model, the results show that usage of social media for information, mode of payment, Rates/ Fare and Classes have a significant effect on the customer satisfaction.

5. CONCLUSION AND RECOMMENDATIONS

According to the study, things are fast changing and this is the same trend with sources of predictive analytics like the social media channels for example which are fast growing and the issue of Big Data analytics comes in handy. The study concluded that: social media site interactivity has a positive relation to customer satisfaction of Madaraka Express among social media users in Kenya. Freebies/discount offers as a factor directly influences Customer satisfaction of SGR transport among social media users in Kenya. Regular updates/postings as a factor has a positive influence on Customer satisfaction with SGR among social media users in Kenya due to customers regular use of online resources to get information relating to the Madaraka express. The results of the study imply that site interactivity, freebies/discounts and regular updates are important elements of Customer satisfaction with Madaraka Express among social media users. The influence of discounts was ranked to have the highest influence on customer satisfaction. The study also clearly shows that if Madaraka express manages to put up an attractive and influential social media sites and other Data sources of predictive analytics will be able to get Big Data that will help them understand their customers better and hence be able to satisfy them. Madaraka Express has established itself as a market leader in SGR transport in Kenya and the utilization of ICT namely Big Data analytics in planning for their services is paramount in the satisfaction their customers.

This study recommends that Madaraka Express should grow social media to be the leading customer service channel since the company is driving data strategy; this means that customers are acquiring data enabled handsets daily and this is good for those who prefer online support which has also been enhanced by the E-gain chat which is the latest addition to the social media channel. This would therefore mean that Madaraka Express would have more support channels for customers. The company should get more influencers on its official sites and pages. Influencers are people who are credible and knowledgeable in specific areas and therefore when they communicate, their information has an impact on the public, and this will therefore reinforce the information shared on the official pages and consequently lead to customer

satisfaction. Madaraka express should handle information gathered by the company from its clients should be handled carefully because such information can be used against the company and /or its clients. So employees need to be very careful with information they are trusted with by the company because it can also be used negatively if it falls in wrong hands.

REFERENCES

Capgemi (2014). *Transforming Insurance Risk Assessment with Big Data: Choosing the Best Path*. Retrieved on June, 2018, from https://www.nl.capgemini.com/resource-fileaccess/resource/pdf/transforming_insurance_risk_assessment_with_big_data_choosin g_best_path_0.pdf

Coker, F. (2014). *Pulse: Understanding the Vital Signs of Your Business* (1st ed.). Bellevue, WA: Ambient Light Publishing.

Dumbill, E. (2012). *Planning for Big Data. A CIO's Handbook to the Changing Data Identifying Patterns in the Chaos*. Rockland, Canada: Syngress Publishing Inc.

Finlay, S. (2014). *Predictive Analytics, Data Mining and Big Data. Myths, Misconceptions and Methods* (1st ed.). Basingstoke: Palgrave Macmillan.

Korn, S. (2011). *The Opportunity for Predictive Analytics in Finance*, HPC Wire

Letouze E. (2011). *Big Data for Development: Challenges & Opportunities*, Global Pulse, New York.

Lindert, B. (2014). *Eckerd Rapid Safety Feedback Bringing Business Intelligence to Child Welfare*. The Huffington Post.

Marz, N. (2015). *Big Data: Principles and Best Practices of scalable real-time data systems*, Manning Publications Company, New York.

Michèle, M. (2010). New Technology Taps 'Predictive Analytics' to Target Travel Recommendations, Travel Market Report, retrieved on September 10, 2015, From which link

Moreira-Matias, L., Gama, J. Ferreira, M., Mendes-Moreira, J. & Damas, L. (2016). Time-evolving O-D matrix estimation using high-speed GPS data streams. *Expert Systems with Applications.* 44: 275 — 288.

Ramachandran, S. & Malladi, S. (2013). *Leveraging unstructured data for improved decision making: A retail banking perspective.* Bangalore, India: Infosys Limited.

Russell, A. M. (2011). *Mining the Social Web: Analyzing Data from Facebook, Twitter,*

LinkedIn, and Other Social Media Sites. Sebastopol, Carolina: O'Reilly Media, Inc.

Schönberger M V. & Cukier K. (2013). *Big Data: A Revolution that Will Transform how We Live, Work and Think.* Houghton Mifflin Harcourt.

Shaunak, D., Maity, A., Goel, V., Shitole, S. & Bhattacharya, A. (2017). Predicting the popularity of instagram posts for a lifestyle magazine using deep learning. 2nd IEEE Conference on Communication Systems, Computing and IT Applications: 174-177.

Siegel, E. (2013). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* (1st ed.). Wiley.

Watson, H. J. (2014). Tutorial: Big Data Analytics: Concepts Technologies and Applications. *Communications of the Association for Information Systems*, 34(65), 1247 — 1268