Predicting Customer Behavior with Combination of Structured and Unstructured Data

Ibukun T. Afolabi, Rowland E. Worlu, Adebayo O. P., Oluranti Jonathan

Email: ibukun.fatudimu@covenantuniversity.edu.ng

Abstract. Presently, there are numerous e-marketing and m-marketing mediums that exist such as YouTube, SMS, What Sapp, Google, twitter, yahoo, Facebook, LinkedIn, email and personal blogs. These mediums are beginning to be used for marketing purposes, particularly by the SMEs in Nigeria. The aim of this research is to address the problem of deciding which of the mediums mentioned above is mostly appropriate to target customer of a particular SME and also to discover the type of data that is most appropriate for analysis in making this decision. In order to achieve this, data was gathered by administering questionnaires and pre-processed based on structured and unstructured data sources. The J48 decision tree classification algorithm was used to mine the data, relevant predictions were made from the structured and unstructured data and the results were evaluated. The results revealed that predicting from unstructured data expresses more of popular opinion, so decision can start from unstructured results and be fined tuned or validated with predicting from structured data. Though structured prediction appears to be better than unstructured, unstructured prediction is still very valuable in situations where there are no structured data such as analysing text messages. Also, Models developed for predicting customer behaviour as regards the marketing channels studied, will form the foundation for marketing decision making, in small and medium businesses in Nigeria.

Keyword: Data Mining, Classification Algorithm, Marketing, e-marketing, m-marketing, Structured data, Unstructured data.

1. Introduction

Customer behaviour research is the study of individuals, groups or organizations in a bid to understand how they select products and also secure, use and dispose products, services, experiences or ideas [1]. According to [2], the need to predict Customer behaviour is of great significance to marketers. This is obvious for the following reasons; first, it is the foundation direct marketing, and secondly, it helps to maximize profit by reducing marketing resources and so much more. Target marketing was defined by [3] as dividing markets into small groups containing the buyers who have unique needs, characteristics, or behaviours. Furthermore, they might require different products or marketing mixes." This of course, is the foundation for successful business intelligence systems. In this study, we focus on predicting customer behaviour from both structured and unstructured data for making quality decisions as regarding marketing small businesses in Nigeria. Data available for analysis as regards the target marketing are in form of both structured and unstructured. Structure data is usually represented in form of table, columns, rows, indexes and so on. It is used to capture transactions, financial reports, word definition and so on. Usually this type of data is characterised by a high degree of predictability [4,5]. Data mining task carried out on structured data can be referred to as the analytic process designed to retrieve consistent patterns and systematic relationship between variables. This is most often validated by deploying the detected patterns to new subsets of data (http://www.statsoft.com/textbook/stdatmin.html #mining). Currently, the amount of information available to the company is mostly unstructured rather than structured. This therefore makes structured mining to be limited in its ability to solve current problems [6,7].

Unstructured data on other hand comes in the form of emails, medical reports, warranties, contracts and so on. They do not have any form of order built into them. Such data have no rules that inform their creation or

usage. Since they are usually in form of text, they do not have key, indexes, columns or attributes [4]. Unstructured data can be stored as excel files, web blogs and so on. The last category of data is the semi-structured data. This type of data is an intermediate between structured and unstructured data. In semi structured data environment, meta-data is usually attached to the data [8]. XML (Extensible Mark-up Language) data storage is an example of semi structured data. Unstructured data mining is referred as text mining. Text mining can be formally defined as extracting interesting and non-trivial patterns of knowledge from unstructured text documents. [9] also defined it as an extension of data mining or knowledge discovery from unstructured databases.
According to [10], SMEs, which is the case study for this research does not have a universally accepted definition, but can be classified into small, medium or large, and this classification definition is determined by different countries. A small business is perceived as one having a total asset in equipment, plant, capital and working capital of less than N250,000 which has 50 full time workers as its employment capacity. The Central Bank of Nigeria [11] defined SMEs as an enterprise hose yearly turnover runs between N25, 000-N50, 000. The Nigerian Industrial Development Bank (NIDB) characterized business as little scale endevour with venture cost (investment and working capital) not surpassing N750,000 while it characterized as medium scale those endeavors whose venture costs falls between the scope of N750,000 to N3 million [12]. According to [13], the Nigeria businesses environment is quite different from the developed economies and other developing countries. It is characterised by challenges such as inadequate infrastructures, such as internet etc. Also there is lack of proper government support programmes for SMEs. To make the matters worse SME operators possess attitudinal problem which greatly contribute to the lack of success of these SMEs. Also [14], underscored that one of the significant marketing issues confronting independent company ventures in Nigeria is absence of adequate comprehension and the use of marketing ideas. Regarding the unique characteristics of SME's in Nigeria as described above, predicting customer behaviour for the purpose of business intelligence becomes very important.

2. Related Works

Broadly speaking, customer behaviour prediction research has been researched from two perspectives. One perspective has been Cognitive-based, which is mostly a theoretical approach i.e through strength of motivation, attitudes, motives, personality traits and learning styles[15]. It also involves the use of questionnaire framework, designed in attempt to understand 'what is going on inside people's heads' [16]. The second perspective is Data mining which deals with the discovery of hidden patterns, predicting future trends and behaviours from data. Data mining over the year has proven to be a more efficient alternative [17]. The implementation methods used in customer behaviour prediction as revealed by the reviewed literature focused mainly on the following category of methods for customer behaviour prediction.
Statistics: the statistical approached using in customer behaviour prediction includes the following; auto regression, logistic regression, linear regression, structural equation modeling and so on. [18-24]

Clustering: The most commonly used clustering technique in customer behaviour prediction, which is, K-means clustering. It works as follows; given an arrangement of points in an Euclidean space and a positive integer k (the number of clusters), K means split the points into k clusters with the goal that the total difference of the (squared Euclidean) distances of each point to its nearest cluster center is limited [25]. This approached is reported to be used for customer behaviour prediction by [18, 26, 27].

Classification: The family of techniques in this category as regards customer behaviour prediction includes; Naïve Bayes [28, 29], Bayesian Network [28, 30], Decision Tree (DT) [31, 32] and Support Vector Machine (SVM) [32, 33].

Psychological prediction Models: This classification involves the group of prediction models which has its foundation in the psychological field of study [34, 35].
Boosting algorithms: The boosting algorithms used for prediction includes Real AdaBoost, Gentle AdaBoost and Modest AdaBoost [36].

Particle Swarm Optimization: prediction method used in customer behaviour is an evolutionary data mining technique [27].
Neuro-Fuzzy: This prediction technique is a combination of neural network and fuzzy logic technique in data mining [32].

Graph Mining Technique: Graph mining is seen as a method of excerpting specific patterns from a combination of graph data [37]. According to [38] it has also been used for customer behaviour prediction.

The remaining technique includes Artificial Neural Network (ANN) [32, 38, 33] and Association Rule Mining [39, 40, 30, 41]. Finally, the dataset used for prediction includes, data gathered from Questionnaire survey and data gathered from Public data repository.

From the literature reviewed, it is discovered that predicting customer behaviour is an on-going and is important to organizations. Most studies researched customer preservation estimation and this was done for the most part utilizing organizational data as dataset for estimating which is for the most part organized. In conclusion, data mining approaches were utilized frequently in predicting customer behaviour, when contrasted with the theoretical approaches. In our research we focus on using a classification technique in data mining to predict customer behaviour.

3. Statement of Problem

Presently, there are numerous e-marketing and m-marketing mediums that exist such as YouTube, SMS, Classification, Google, twitter, Yahoo, Facebook, LinkedIn, Email and Personal blogs. These mediums are beginning to be used for marketing purposes, particularly by the SMEs in Nigeria. This is mostly due to the fact that this type of business does not have enough capital to invest in marketing [14] through mediums like media (TV stations, Radio stations). E-marketing and m-marketing mediums provide a cheaper means of marketing to their customers. There is therefore a problem in Customer Relationship Management, particularly relating to marketing which has to do with deciding which one of the mediums mentioned above is most appropriate to target customer of a particular SME. This problem is not trivial because it leads to wasting of resources (financial, time etc.) used in marketing through these mediums. Also according to [46] 'Reaching out to the target audience' is the biggest challenge marketers faced in email marketing for example. The aim of this research is therefore, to propose customer behaviour prediction models in electronic marketing (i-marketing and m-marketing) for the Nigerian context. A typical problem scenario to illustrate this problem is to attempt to decide which marketing avenue is able to target and retain a potential customer of a table water production business. The company that produces this table water may be deposed to certain information about the potential customer such as age, computer literacy, average income etc. This company may need to decide as to which channel can reach a particular target customer out of goggle marketing, YouTube marketing etc. This decision becomes non trivial, especially if they don't have the resources to marketing through all the available mediums.

4. Methodology

The methodology used in this research is based on the following major steps:

i. Gather data for predicting customer behaviour
ii. Pre-process the data based on structured unstructured basis
iii. Predict from structured and unstructured data
iv. Evaluate & deploy the model to predict on fresh data.

4.1 Data Gathering

The survey instrument was employed to collect data by administering 400 questionnaires whereas 348 were gathered from the respondents. Both online questionnaire and paper-based questionnaire survey were administered. The questionnaire consists of two sections so as to bring out complete characteristics of the respondent demographic analysis. Respondents personal information was gathered in the first section, information such as age, gender and education background. While respondent's perception about electronic marketing was gathered in the second section.

The questionnaires contained sections for both structured and unstructured questions. Structured questions allowed the respondents to select from alternatives given while the unstructured provided a space for respondents to express themselves in writing. The structured part of the questionnaire retrieved the following formation from the respondents: age, gender, marital status, occupation, highest academic qualification, states and town the respondents reside in Nigeria, level of computer literacy, average income per month, hobbies, income range, tribe, religion, internet subscription per month and frequency of your visiting YouTube; Classification; Facebook; Twitters; Google and Personal blogs and so on.

The unstructured part of the questionnaire allowed the respondents to express themselves freely, thereby retrieving information concerning their interest in life, their attraction to online business, their dislike about online business, attraction and dislike about doing business through the mobile phone platform and so on.

4.2 Data Pre-processing

The data pre-processing stage consisted of two types, structured pre-processing and unstructured pre-processing. For structured pre-processing, the data was converted to nominal data. Data in the unstructured part of the questionnaire was filtered to remove words that were termed not to be important from documents content. Words that are characterized as conjunctions and prepositions, determiners, pronouns, articles, non-informative verbs and common verbs. Due to this process, highly relevant or important words are selected. After this, the extracted term is stemmed – a method of removing word prefixes an suffixes (such as merging both responsibilities and responsible to responsible). Finally, TF-IDF (Term Frequency, Inverse Document Frequency) – a weighting algorithm - is used to allocate weight values to words to differentiate them syntactically in a document [42]. All the above was carried out using the weka toolkit [43].

4.3 Customer Behaviour Prediction

In order to predict from structured and unstructured data i.e. generate model/classifier that is able to predict on fresh data, the J48 algorithm is used. Weka (Waikato Environment for Knowledge Analysis) data mining tool [43] was used to pre-process the data and implement the classification algorithm. The particular algorithm used was the j48 classification algorithm. The data gathered were divided into two. The first contained two thirds of the data which was used to develop the classifier and the set of data remaining was used to test and evaluate the classifier model developed.

By definition, a classification algorithm builds a model of classes from a set of records that contain class labels.

Decision Tree Algorithm is a classification algorithm that finds out how the attributes-vector behaves for a number of instances [44]. J48 is an extension of ID3, which has additional features of accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. J48 is an open source Java implementation in WEKA and is made up of the C4.5 algorithm. The following is the three steps involved in the algorithm;

- Instances that belong to the same class a leaf represents the tree, therefore, the leaf is identified with the same class.
- The possible information is computed for every attribute, given by a test on the attribute. Then the gain in information is computed that would result from a test on the attribute.
- Then the best attribute is found on the basis of the present selection criterion and that attribute selected for branching [45].

4.4 Evaluation and Prediction

To deploy and make predications with the model, fresh data were gathered from 10 respondents, so as to make predictions of their behaviour as regards these marketing channels. The predicted result was now compared with the true decisions on of these respondents on the marketing avenues to evaluate the success rate of the models used for prediction.

5. Results of the study

5.1 Model Evaluation Results

The data gathered were divided into two for both the structured and the unstructured part of the questionnaire. The first contained two thirds of the data which was used to develop the classifier/model and the set of data remaining was used to test and evaluate the classifier model developed. Table I displays the model evaluation results for both the structured and unstructured data.

Table1. Model Evaluation Results

Model	Evaluated Results(correctly classified Instances)-Structured Data	Evaluated Results (correctly classified Instances) -Unstructured Data
EmailPredictionModel	77.1429 %	79.1209 %
FacebookPrediction Model	85.7143 %	68.1319 %
GooglePredictionModel	93.5484 %	81.3187 %
PersonalBlogsPredictionModel	72.8261 %	74.7253 %
SMSPredictionModel	91.8919 %	68.0000%
TwitterPredicitonModel	73.913 %	70.3297 %
WhatsappPrediction Model	84.7826 %	73.6264 %
YoutubePrediction Model	70.3297 %	69.2308 %

Table 2a and 2b show the predicted results for the fresh data as for structured data. In Table 2a, 2b,4a and 4b "none" represents that the individual will not decide for any of the option, "VeryFreq" means that the individual will visit the marketing medium in question very frequently, "NeverVis" means that the individual will never visit the marketing medium, "Frequent" means that the individual will visit the medium frequently, "NotFrequ" means that the individual dose not visit the medium frequently, "NeverVis" means that that individual has never visited the medium while "No Prediction means that the model could not predict for that particular individual. "Yes" indicates a person that has done business or has responded to adverts via the SMS platform while "no" indicates the opposite. "Never" means never does business as a result of email marketing, "often" means often does, "Notoften" means not often do business as a result of email marketing.

Table 2a. Structured Mining Prediction Result Part1

Person-ID	Email classifica-tion output	Facebook classification output	PersonalBlog classification output	youtube classification output
1	none	VeryFreq	none	none
2	never	NeverVis	NeverVis	NotFrequ
3	never	VeryFreq	none	none
4	often	VeryFreq	Frequent	NotFrequ
5	often	VeryFreq	VeryFreq	NotFrequ
6	never	VeryFreq	NeverVis	NeverVis
7	Not often	VeryFreq	Frequent	No Prediction
8	Not often	VeryFreq	none	No Prediction
9	often	VeryFreq	VeryFreq	No Prediction
10	often	VeryFreq	none	No Prediction

From Table 2a and 2b, it is obvious that Facebook, Google and Classification is the most visited by the individuals used for the experiment and YouTube is the least visited. A lot of things could be responsible for this, it could be attributed to the fact that the kind people investigated are low income people who do not spend much on internet facility and therefore can't afford to be watching or downloading videos on the internet. Whatever is the reason for this, it makes it clear that trying to target these set of people via YouTube will be a waste of resources.

Table 2b. Structured Mining Prediction Result Part2

Person ID	Google classification output	twitter classification output	SMS classification output	whatsapp classification output
1	None	none	No	none
2	VeryFreq	NotFrequ	No	VeryFreq
3	VeryFreq	none	No	VeryFreq
4	VeryFreq	Frequent	Yes	VeryFreq
5	VeryFreq	VeryFreq	Yes	VeryFreq
6	VeryFreq	NeverVis	No	VeryFreq
7	VeryFreq	VeryFreq	Yes	VeryFreq
8	VeryFreq	none	Yes	VeryFreq
9	VeryFreq	VeryFreq	Yes	VeryFreq
10	VeryFreq	VeryFreq	Yes	VeryFreq

Table 3a. Evaluation Result for Structured Mining Part 1

PersonID	Email Prediction evaluation	Facebook Prediction evaluation	PersonalBlogs Prediction evaluation	youtube Prediction evaluation
1	Correct	Correct	Correct	Correct
2	Correct	Correct	Correct	Not correct
3	Correct	Correct	Not correct	Correct
4	Correct	Correct	Correct	Correct
5	Correct	Correct	Correct	Not correct
6	Correct	Correct	Correct	Correct
7	Correct	Correct	Correct	Not Evaluated
8	Correct	Correct	Correct	Not Evaluated
9	Not correct	Correct	Correct	Not Evaluated
10	Correct	Not correct	Not correct	Not Evaluated
% of Correctly Predicted	90%	90%	80%	67%

Table 3a and 3b is the result gotten after comparing the predicted output for each of the mediums by the model developed for each of these mediums and the actual response of the investigated persons. Google, SMS and Classification, gave 100% prediction success rate meaning that the results generated by the models can be relied on 100%. The models for predicting emails and book did not do badly either, they gave 90% success rate. At any rate, none of the models performed less than 50%.

Table 3b. Evaluation Results for Structured Mining Part 2

PersonID	Google Prediction evaluation	twitter Prediction evaluation	SMS Prediction evaluation	whatsapp Prediction evaluation
1	Correct	Correct	correct	correct
2	Correct	Correct	correct	correct
3	Correct	Not correct	correct	correct
4	Correct	Correct	correct	correct
5	Correct	Correct	correct	correct
6	Correct	Not correct	correct	correct
7	Correct	Correct	correct	correct
8	Correct	Correct	correct	correct
9	Correct	Correct	correct	correct
10	Correct	Not correct	correct	correct
% of Correctly Predicted	100%	70%	100%	100%

Table 4a. Unstructured Mining Prediction Result Part 1

PersonID	Email classification output	Facebook classification output	PersonalBlog classification output	youtube classification output
1	NotOften	VeryFreq	Frequent	NotFrequ
2	NotOften	VeryFreq	VeryFreq	NotFrequ
3	NotOften	VeryFreq	VeryFreq	NotFrequ
4	NotOften	VeryFreq	VeryFreq	NotFrequ
5	NotOften	VeryFreq	Frequent	NotFrequ
6	NotOften	VeryFreq	NeverVis	NotFrequ
7	NotOften	VeryFreq	NeverVis	NotFrequ
8	NotOften	VeryFreq	VeryFreq	NotFrequ
9	NotOften	VeryFreq	VeryFreq	NotFrequ
10	NotOften	VeryFreq	NeverVis	NotFrequ

From Table 4a and 4b, it is obvious that Facebook, Google and classification is still the most visited by the individuals used for the experiment and YouTube is the least visited. This result is the similar and almost the same with that of the structured mining prediction. The advantage of this is that, if structured data is not available to carry out this prediction unstructured data which is the most available data in organizations will also be able to give the same inferences

Table 4b. Unstructured Mining Prediction Results Part 2

PersonID	Google classification output	twitter classification output	SMS classification output	whatsapp classification output
1	VeryFreq	NeverVis	Never	VeryFreq
2	VeryFreq	NeverVis	NotOften	VeryFreq
3	VeryFreq	VeryFreq	Never	VeryFreq
4	VeryFreq	NeverVis	Never	VeryFreq
5	VeryFreq	VeryFreq	Never	VeryFreq
6	VeryFreq	NeverVis	Never	VeryFreq
7	VeryFreq	VeryFreq	Never	VeryFreq
8	VeryFreq	Frequent	NotOften	Frequent
9	VeryFreq	VeryFreq	Never	NeverVis
10	VeryFreq	VeryFreq	Never	VeryFreq

Table 5a. Evaluation Results for Unstructured Mining Part 1

PersonID	Email Prediction evaluation	Facebook Prediction evaluation	PersonalBlogs Prediction evaluation	youtube Prediction evaluation
1	correct	Correct	Correct	Correct
2	correct	Not Correct	Not Correct	Not Correct
3	correct	Correct	Correct	Correct
4	Not Correct	Correct	Correct	Correct
5	Not Correct	Correct	Correct	Not Correct
6	correct	Correct	Correct	Correct
7	correct	Correct	Not Correct	Not Correct
8	correct	Correct	Correct	Correct
9	correct	Correct	Correct	Not Correct
10	Not Correct	Not Correct	Not Correct	Correct
% of Correctly Predicted	70%	80%	70%	60%

Table 5a and 5b is the result gotten after comparing the predicted output for each of the mediums by the model developed for each of these mediums and the actual response of the investigated persons for the unstructured data. only google gave 100% prediction success rate. Classification and Facebook gave 80% each. in the unstructured mining evaluation results, the model that predicted for YouTube and SMS marketing produced 60% and 40% success rate.

Table 5b. Evaluation Results for Unstructured Mining Part 2

PersonID	Google Prediction evaluation	Twitter Prediction evaluation	SMS Prediction evaluation	whatsapp Prediction evaluation
1	Correct	Correct	Correct	Correct
2	Correct	Correct	Correct	Non Correct
3	Correct	Correct	Correct	Correct
4	Correct	Non Correct	Non Correct	Correct
5	Correct	Correct	Non Correct	Correct
6	Correct	Non Correct	Correct	Correct
7	Correct	Correct	Non Correct	Correct
8	Correct	Correct	Non Correct	Correct
9	Correct	Correct	Non Correct	Non Correct
10	Correct	Non Correct	Non Correct	Correct
% of	100%	70%	40%	80%

Table 6. Comparing Structured and Unstructured Evaluation

Marketing Channels	Structured Prediction Evaluation	Unstructured Prediction Evaluation
Google	100	100
Twitter	70	70
SMS	100	40
Whatsapp	100	80
Email	90	70
Facebook	90	80
PersonalBlogs	80	70
Youtube	67	60

Comparing the result of the evaluation of prediction results for structured and unstructured data as expressed in Table 6, the following observation and recommendations are made:

1. Structured data predication model does better than unstructured data predication model but they both point towards the same direction.
2. It was discovered that predicating from unstructured data expresses more of popular opinion, so decision can start from unstructured results and be fined tuned or validated with predicting from structured data.

3. Even though structured prediction appears to be better than unstructured, unstructured prediction is still very valuable in situations where there are no structured data such as analysing text messages etc.

6. Conclusion and Further Studies

In conclusion, the models developed for predicting customer behaviour as regards the marketing channels studied will form the foundation for marketing decision making in small and medium businesses. Such models can be embedded in decision supports systems for marketing purposes. For example revisiting the problem scenario, we can see from the result, for PersonID no "2" for example in the structured Prediction, that the problem of deciding which marketing avenue is most appropriate for a target marketing campaign is solved by choosing between Google and classification. Surprising as this maybe, such a potential customer is not a fan of Facebook. Also, this study is not trivial because it will help to achieve one of the main goals of CRM (Customer Relationship Management), which is to turn a target customer to a paying customer and a lifetime customer and eventually a marketing customer. Involving big data analysis is planned to obtain an all-inclusive source of the unstructured data.

Acknowledgement
The authors of this research do appreciate the Covenant University Centre for Research, Innovation, and Discovery (CUCRID) for its support for this research.

References

[1] Raorane A. & Kulkarni R.V, (2011) "Data Mining Techniques: A Source for Consumer Behavior Analysis," Shahu *Institute of business Education and Research, 1-15*.
[2] Durmaz Y. and Diyarbakırlıoğlu I., (2011). "A Theoritical Approach to the Strength of Motivation in Customer Behavior" in *Global Journal of Human Social Science 11(10) 1.0*
[3] Sulekha G. (2011) "The basis of market segmentation: a critical review of Literature," *European Journal of Business and Management ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online) 3(9), 2011 45-55*
[4] Inmon, B. (2007) "Structured and Unstructured Data, Bridging the gap," in Business Intelligence Network's Bill Inmon Channel. http://www.b-eye-network.com/view/4955
[5] Worlu, Rowland E.K. (2010) *Marketing Strategies of Nigerian Political Parties: A Comparative Analysis.* Global Journal of Management and Business Research, 10 (5). pp. 48-63.
[6] Unitas Corporation, (2002) "A Single View: Integrating Structured and Unstructured Data/Information with the Enterprise," in unitas, the portal is the business ™,. http://lsdis.cs.uga.edu/GlobalInfoSys/Structured-and-Unstructured-for-EIPs.pdf.
[7] Blumberg, R., and Atre, S.(2003) "The Problem with Unstructured Data," DM Review 13(4),
[8] Ukelson, J. (2006) "Combining Structured, Semi structured and Unstructured Data in Business applications," in DM Direct Newsletter.
[9] Ah-Hwee T., "Text mining: The state of the art and the challenges", 2006.

[10] Gulani M. G. and Usman A., (2012). "Financing Small and Medium Scale Enterprises (SMEs): A Challenge for Entrepreneurial Development in Gombe State" in *Asian Journal of Business and Management Sciences ISSN: 2047-2528 2 (9) 17-23*

[11] CBN, (2011). Micro Finance Policy, Regulatory and Supervisory Framework for Nigeria. A Publication of Central Bank of Nigeria, Abuja, Nigeria

[12] Ogbadu E. E., (2012). "Appraisal of the practical application of marketing research by SMEs in Nigeria" Kuwait Chapter of *Arabian Journal of Business and Management Review 2(2)*

[13] Ogundele O. J. K., Akingbade W. A., SakaR. O.,ElegundeA. F. and AliuA. A, (2013). "Marketing Practice of Small and Medium Enterprises (SMEs): Perspective from a Developing Country" in *Mediterranean Journal of Social Sciences MCSER Publishing, Rome-Italy 4(3)*

[14] Ogechukwu A. D., Oboreh J. S., Umukoro. F and Uche A. V., (2013) "Small and Medium Scale Enterprises (SMEs) in Nigeria the Marketing Interface" in Global *Journal of Management and Business Research Marketing 13(9) 1*

[15] Holdershaw J., and Gendall P., (2008). "Understanding and predicting human behaviour" in ANZCA08 Conference, *Power and Place*. Wellington

[16] Yada, K., Motoda, H., Washio, T. and Miyawaki, A. (2005) Consumer behaviour analysis by graph mining technique, New Mathematics and Natural Computation, 2 (1), 59-68.

[17] Rajagopal S., (2011). "Customer Data Clustering Using Data Mining Technique" in *International Journal of Database Management Systems (IJDMS) 3(4)*

[18] Vijayalakshmi S., Mahalakshmi V., and Magesh S., (2013). "Knowledge discovery from consumer behavior in electronic home appliances market in Chennai by using data mining techniques" in *African Journal of Business Management 7(34), 3332- 3342*

[19] Khan I., (2012). "Impact of Customers Satisfaction And Customers Retention on Customer Loyalty" in *International Journal of Scientific & Technology Research 1(2)*

[20] Sadasivan K., Rajakumar S., &Raijnikanth, R., (2011). "Role of Involvement and Loyalty in Predicting Buyer's Purchase Intention towards Private Apparel Brand Extensions" in *International Journal of Innovation, Management and Technology, 2(6), 519-524.*

[21] Barrios, E. B. and Lansangan J.R.G., (2012). "Forecasting Customer Lifetime Value: A Statistical Approach" in *Philippine Management Review,* 19

[22] Banerjee S., and Pawar S., (2013). "Predicting Consumer Purchase Intention: A Discriminant Analysis Approach" in NMIMS Management Review 23 129

[23] Sandy, C. J., Gosling, S. D., and Durant, J. (2013). "Predicting Consumer Behavior and Media Preferences: The Comparative Validity of Personality Traits and Demographic Variables. Psychology & Marketing

[24]Dirisu, Joy Favour and Worlu, Rowland E.K. and Osibanjo, Adewale Omotayo and Borisade, Taiye and Olokundun, Ayodele Maxwell and Atolagbe, Tolulope Morenike, and OBI, JAMES NWOYE (2018) *Dataset on brand culture and perceived value of offerings to customers in the hospitality industry in Nigeria.* Data in Brief, 19. pp. 1-5.

[25] Lloyd S.P. (1982) Least squares quantization in PCM. Unpublished Bell Lab. Tech. Note, portions presented at the Institute of Mathematical Statistics Meeting Atlantic City, NJ, September 1957. Also, IEEE Trans Inform Theory (Special Issue on Quantization), IT-28, 129-137

[26] Hosseini S. M., Maleki A., and Gholamian M. R., (2010). "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty" in Expert Systems with Applications 37, 5259–5264

[27] Liu J. L. and Chen C. L., (2012). "Application of Evolutionary Data Mining Algorithms to Insurance Fraud Prediction" in Proceedings of 2012 4th International Conference on Machine Learning and Computing IPCSIT 25

[28] Kirui C., Hong L., Cheruiyo W., and Kirui H., (2013). "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining" in *IJCSI International Journal of Computer Science Issues, 10(2) 1*

[29] Mauser A., Bezrukov I., Deselaers T. and Keysers D.(2005) Predicting Customer, Behavior using Naive Bayes and Maximum Entropy – Winning the Data-Mining-Cup, 2004. Proc. Informatiktage 2005 der Gesellschaftf¨urInformatik, in press, St. Augustin, Germany.

[30] Qiu, Jiangtao, (2014)"A Predictive Model For Customer Purchase Behaviour In E-commerce Context" (2014). *PACIS 2014 Proceedings.* Paper 369

[31] D'HaenJ., Van den Poel D., and Thorleuchter D., (2013). "Predicting customer profitability during acquisition: Finding the optimal combination of data source and data mining technique" in Expert Systems with Applications 40 2007–2012

[32] Abbasimehr H., Setak M., and Tarokh M., (2014). "A Comparative Assessment of the Performance of Ensemble Learning in Customer Churn Prediction" in *The International Arab Journal of Information Technology, 11(6)*

[33] Zhang Y., and Zhao Z., (2014). "Study on Consumer Behavior Predict in E-commerce Based on Multi-Agent" in *International Journal of u- and e- Service, Science and Technology 7(6) 403-412*

[34] Taubinsky, D. (2013), `From intentions to actions: A model and experimental evidence of inattentive choice', working paper

[35] Ibrahim Y., and Vignali C., (2005). "Predicting Consumer Patronage Behaviour in the Egyptian Fast Food Business" in Innovative Marketing 1(2)

[36] Shao Jinbo, Li Xiu, and Liu Wenhuang (2007), "The Application ofAdaBoost in Customer Churn Prediction," In *Proceedings of Service Systems and Service Management, 2007 International Conference* 1-6

[37] Inokuchi, A., Washio, T., Nishimura, Y., Motoda, H. (2002) General Framework for Mining Frequent Structures in Graphs. Proc. of the International Workshop on Active Mining. 23-30.

[38] Sharma A., and Panigrahi P. K., (2011). "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services" *in International Journal of Computer Applications 27(11) (0975 – 8887)*

[39] Haastrup A. V., Oladosu O. A., Okikiola, F. M, Oladiboye O. E., and Ishola P. E., (2014). "Customer behaviour analytics and data mining" in *American Journal of Computation, Communication and Control 1(4): 66-74*

[40] Zheng B., Thompson K., Lam S.S., and Yoon W. S., (2012). "Customers' Behaviour Prediction Using Artificial Neural Network" in Proceedings of the 2013 Industrial and Systems Engineering Research Conference A. Krishnamurthy and W.K.V. Chan, eds
http://www.statsoft.com/textbook/stdatmin.html

[41] Kerdprasop N., Kongchai P., and Kerdprasop K., (2013). " Constraint Mining in Business Intelligence: A Case Study of Customer Churn Prediction" in *International Journal of Multimedia and Ubiquitous Engineering 8(3)*

[42] Hany M.; Dietmar R.; Nabil I. and Fawzy T. (2007) "A Text Mining Technique Using Association Rules Extraction", *International Journal of Computational Intelligence 4(1) 2007 ISSN 1304-2386.*

[43] Hall M. Eibe Frank, Holmes G., Pfahringer B. (2008). The WEKA Data Mining Software: An Update in SIGKDD Explorations 11(1) 10-13

[44] Nadali, A; Kakhky, E.N.; Nosratabadi, H.E.,(2011) "Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system," Electronics Computer Technology (ICECT), 3rd International Conference on , 6, 161-165.

[45] Korting, Thales Sehn. "C4. 5 algorithm and Multivariate Decision Trees." Image Processing Division, *National Institute for Space Research--INPE.*

[46] Octane Research (2015) Introducing e-Marketing Outlook for India (http://octaneresearch.in/wp-content/uploads/2015/01/The-Digital-DNA-The-State-of-Emarketing-in-India.pdf)