An Application of Association Rule Learning in Recommender Systems for e-Commerce and its Effect on Marketing

Anne W. Mbugua
Faculty of Information Technology
Strathmore University
Nairobi, Kenya
annewangari18[at]gmail.com

Allan O. Omondi
Faculty of Information Technology
Strathmore University
Nairobi, Kenya
allan[at]odhiambo.me.ke

***Abstract*—High annual customer churn rates and low customer attractions caused by poor marketing recommendations inhibit enterprises from making as much profit as they should. The purpose of this research was to derive a more optimized association rule learning algorithm that can be used in a web-based recommender system for small-scale enterprises. The method used was a case study approach on a small-scale enterprise called Makewa Hardware located in Ruiru, Kenya. Having access to the enterprise supported the use of the agile methodology, more specifically, extreme programming in the development of the system that applied the algorithm. A sample of training data consisting of transactions made in the past was obtained from the enterprise in order to create the machine learning aspects of the algorithm. The results showed that the derived association rule learning algorithm was able to learn and generate its own frequent-item-set and use this to give appropriate recommendations to customers. The results revealed the system's ability to make more accurate recommendations. This was based on the pattern of purchases made from the hardware store by various customers. The recommendations were given on a weekly basis. The implication of the results on the subjects showed that more business owners are open to having intelligent systems help make and predict their sales. The findings can be applied not only in hardware stores but also in other retail stores. Future research can ensure that a normal dataset can be transformed into a market basket without it losing important information.**

Keywords— Association rule learning; e-commerce; marketing; recommender systems

I. INTRODUCTION

In this day and age where repairs and installations no longer need handymen to do all the work, starting up a hardware store could be one of the most profitable business ideas available. This however, as insignificant as it may seem is a grand and worthwhile venture and good planning would be the gap between riches and defeat.

Research by [1] says that before opening a business, it is key to investigate your area to determine its maturity in the future, your competitors' weak points and the purchasing power of your customers. However, when starting the business, these were not factors that the business owner put into consideration. Her main aim was to begin a business and sustain her family's basic necessities. This has not brought adequate success even with the objectives she originally had. The business itself is in a remote area where not many customers are able to access easily. When setting up a hardware business, the entrepreneur needs to determine how he or she can compete with larger markets and not get terrified by their position in the market. Research shows that small scale hardware stores like Makewa hardware, are in a position to efficiently and effectively compete in the market and also, many customers prefer individualistic retail outlets for the high quality of services they get.

Research by [2] showed that 11% of entrepreneurs experience a challenge when it comes to devising the most appropriate marketing techniques for their business; and this problem is no exception for Makewa Hardware. The research acknowledges that there are many more solutions to this problem. For example, the owner of the business may decide to change the location of the business or she could decide to partner with a larger organization in the market. However, this could come with its own limitations such as, her losing most of the shareholdings in her own business. For this reason and more, this research submits this solution as the one to be implemented in the business.

A poor marketing strategy is a challenge faced by Makewa hardware because many customers do not know the business exists and the ones who do, don't know about the products existing in the hardware store. The strategies used to market the products is by word of mouth and a sign post at the entrance of the door. Makewa's sales are increasingly dwindling as a result of a decline in the number of customers due to factors like migration to more developed areas.

We propose the development of a web based system that would set the business apart from its competitors due to the quality services offered with the integration of technology in

the business thus increasing its profits and earning it a share in the market. This solution has a two-pronged approach to it. First, the web-based approach. This aims to inform potential customers that the business exists. Second is a recommender system that uses association rule learning to discover interesting relations between hardware products. This is done with the aim of answering the research question of this paper which is: "How can an intelligent e-Commerce system positively affect the performance of such businesses?"

The remainder of the paper is organized as follows. Section II presents the methodology applied in the research. This includes software development life cycle (SDLC) methodology, as well as a description of the research site and the sample analyzed. Section III starts by introducing the reader to the technology involved and builds up on this introduction in order to present the results of the research. Finally, Section IV presents a discussion to conclude the paper.

II. METHODOLOGY

An agile methodology, specifically extreme programming was applied in this research. We chose this method because the agile methodology factors in the need to have a system that is flexible enough to cater for the ever changing needs of the customer. This research also factors in the analysis of the result of manipulating the independent variable through the use of treatments. In this research, recommender systems for e-Commerce (the dependent variable) depends on association rule learning (the independent variable) which is acted upon by algorithms such as apriori, eclat, fp-growth algorithm, and the derived hybrid algorithm.

A. Research Site

A particular case study was taken from Makewa Hardware Store, a business that is situated in Ruiru, Kenya. This business deals in selling hardware products to customers who could be residents, local construction workers or aspiring homeowners. Between the two options available for starting a hardware business, which are entrepreneurship and purchasing a hardware franchise, this particular business was opened by the owner herself. It enabled her to choose the items she needed in her store unlike having a franchise which limits you to the products it specifies.

Makewa hardware, however, does not receive as many customers because of two main reasons: The remote setting of the business and the poor marketing strategy used. The marketing strategy applied by the business is word of mouth and a sign post at the entrance of the shop. The poor marketing strategy leads to two necessities. Firstly, there is need to inform new customers that the business exists and second, there is need to inform the current customers that the shop has certain products on sale. Due to this, the sales are low throughout the year. This leads to the problem that this project proposes to address. The problem is the fact that the business is not making as much profit as it should because of high annual customer churn rate and low customer attraction caused by a poor marketing strategy.

B. Sample

We collected data from a sample of sales from Makewa Hardware Store with personalized information. The objective of collecting this data was to identify the relationship that exists between the products a customer purchases from the hardware store.

TABLE I. SAMPLE CUSTOMER PURCHASE

Date	Customer	Quantity	Item	Price (Ksh)
22/07/2015	Customer X	1-2.5KG	Polyfilla	250
		1 pc	Brush	200
		1 pc	Scrapper	50
		1-20 ltr	Vesta White	2,400
		1 pc	Sand paper	100
Total:				**3,000**

To derive the model the application would use, it was necessary to collect data from the hardware store as would serve as the basis for the recommendations given as shown in Table I. To analyze the collected data and guarantee its consistency and completeness, association rule learning was applied to discover the interesting relations that exist between the items in a customer's item basket.

III. RESULTS

A. Challenges facing businesses today

For most people, starting a business is the only way in which they are able to sustain their family's basic needs. However, in as much as these startups seem to be lucrative, most people go into these business without having analyzed what it means to begin and grow the business. Some of the challenges faced by these SMEs include customer attraction, customer relations and product branding.

Work by [3] states that there are three stages in which customer relationship with a company is represented. The first stage is "attracting customers" stage where the customer has formed attitudes to the brand and product before making purchases. The second is "after the purchase" stage that forms the evaluation of the benefit of the brand and the benefit of the item offer for the customer and as a result the customer continues to purchase from this company. The third stage is the analysis of the established relationship between the organization and the customers.

Assessing the performance of a business has proved to be a difficult task for most business owners. [4] define business performance as the effectiveness of an organization in fulfilling its purpose. The goal of business performance measurement is an improvement in financial returns of an organization. Customer relationship management as [5] put it says it begins with customer identification and majorly involves examining customers who are leaving and how they can be gained back and retained. With aim of improving business performance, [4]

give one of the elements of customer attraction as direct marketing. [6] defines direct marketing as an advancement process that inspires buyers to place orders through mediums.

Retention of customers is also a major problem faced by many companies. [7] defines customer retention as the marketing goal of keeping one's customers from going to the competitors. The ability to retain customers affects both revenues and costs when profitability is measured, such that profit is given as revenue less costs.

B. Recommender Systems

Recommender systems are systems developed for customers and their merchants with the aim of realizing two crucial benefits. To the users, they receive help through the recommendations given by the system to make sufficient purchases, while the business in turn benefits from the recommendations by luring customers to purchase more of their products, thus bringing more profits to the business through the increasing sales.

C. How Recommender Systems Work

Recommender systems study databases of how customers interact with websites and web applications. Results from the study by [8] suggest that the data collected from analyzing the databases are many at times in the form of contents of shopping baskets, explicit rating data and purchase information.

For any recommender system, there are two entities that must be present. These are the customer and the product. The work by [9] indicates that there are three categories of input into and two categories of output from a recommender system. The first input are ratings, otherwise known as votes, which show the opinion of users on products offered on sale. The second input is the demographic data which is data picked specifically from the customer and indicates the age, gender or education. Lastly, is content data which is based on the analysis related to documents rated by a user. The output given by these systems are two: prediction, which represents the expected judgment of the customer; and recommendations, which represents the item a user has the highest chance of taking interest in.

D. Algorithms Applied to Recommender Systems

There are different algorithms that can be applied in recommender systems. They include memory based algorithms that work through the entire user-item matrix to give the prognosis and are mostly categorized under the collaborative filtering category. Work by [9] indicates that the major steps involved in this method are: recommendation generation, representation and neighborhood formation. This research, however, was mainly interested with recommendation generation. It is further classified into two broad categories according to [9]. They are: association rule-based recommendations and most-frequent item recommendations.

Association rule-based recommendations is the main area of study in this research. Studies by [9] show that, between two sets of items, I_X and I_Y , such that $I_X, I_Y \subseteq I$ and $I_X \cap I_Y = \emptyset$, an association rule would state that if items from set I_X appear in transaction T, then there is a high possibility that items from set I_Y would also be found in T. The nature of these association rules is usually gauged by calculating their support and confidence. Research from [10] states the few, but important concepts applied in association rule learning. They include: support, confidence, lift and conviction. All these concepts play a major role in machine learning. Support is an indication of how frequently an item-set such as {brush, paint} appears in the database.

Confidence on the other hand, is an indication of how often the rule has been found to be true. For example, the rule {paintbrush, roller cover} => {paint} would have a confidence of 1.0, meaning that, for all of the transactions, that is to say 100%, containing paintbrush and roller cover, the rule is correct. This is because many of the times a customer buys a paintbrush and a roller cover, paint is bought as well; "many" in this context referring to 100% of the times a customer makes a purchase. Lift, is also one of the concepts applied and refers to the correlation of the noted support to an item-set that is expected if the values X and Y are self-sufficient. Conviction is the ratio of the rate of occurrence in which the rule makes a wrong prediction or the expected frequency that X occurs without Y. Association rules with positive support are key because they report an abundant quantity of products, while association rules with positive confidence. They are key since their prognosis of the end result is correct.

E. Gaps in Existing Algorithms for Recommender Systems

The research we carried out revealed some setbacks with the existing algorithms which include Apriori algorithm, Eclat algorithm and FP-growth algorithm that are meant for Recommender systems. Below are our findings:

(i) The algorithms based on support and confidence may fail to spot some less frequent, but sensible interesting rules.

(ii) Each of the algorithms seem to use what the user has fed into it to make recommendations and with no prior domain knowledge, making it difficult for the user parameters, which are dependent on the data set, to be estimated appropriately.

(iii) The research also did not come across an algorithm that can transform a normal data set into a market basket data set without it losing important information.

(iv) The efficiency of these algorithms is poor when handling large data sets, because of the time and resources needed such as memory.

In order to close the gap between the shortcomings of the three algorithms, the research made use of the information provided by the owner of the hardware store as prior knowledge that the system will use to make better recommendations.

F. Hybrid Algorithm

The hybrid algorithm is the result of this research and an improvement of the existing algorithms. This is because the new algorithm highly sensitizes on the use of prior knowledge from a business. The study done on Makewa Hardware relied

on transactional data from the business which was adequately provided by the owner of the store. The data provided acted as the prior knowledge to be acted upon by the algorithm and provide better recommendations.

This algorithm works in such a way that, the items a customer has in his or her shopping cart are fetched and then the current page item is checked whether it is in the cart. If true, the algorithm proceeds to obtaining all the purchased products and isolating the current cart items. If false, the item is added to the current cart items and proceeds. When true, the algorithm checks to see if the product has been traversed in the dataset and if it has, obtain the current item and give it weight. Afterwards, the products are grouped in frequencies and cumulative weights of occurrence are assigned accordingly. The products are then sorted in descending order and the top four items are obtained, which are then given as recommendations to the item selected by a customer.

The reason why the recommendations relied on prior knowledge from the business's transactional data, was to have the algorithm learn the purchasing patterns of the customers in the area. This was done with aim of giving more relevant recommendations to a customer who might be visiting the site for the first time or could be a regular customer.

IV. Discussion

This research confirms that small and micro-entrepreneurial businesses experience quite a number of drawbacks, especially in marketing of their products. Being in the 21st century, withholding the adoption of Information Technology in businesses, is a serious mistake because it leaves a business at the mercy of its competitors and highly limits the success of a business. Sustainable economic development as stated by [11] arises from the potential to create innovations in business processes to gain an aggressive advantage that is included in terms of better management and raises the chances of business success.

With the introduction of IT in a business, businesses can now achieve more success owing to the fact that this is a new era; the digital era. However, it is only through the use of an efficient information system that a business can fully achieve its set objectives. Recommender systems are an example of such systems that fully meet the needs of a business, set to promote sales and still help a customer find the products they are looking for. The results show that an efficient and effective algorithm, such as the derived hybrid algorithm, if implemented in a business, is capable of bringing success just as it has brought to Makewa Hardware. Not only can recommender systems be used in hardware stores, they can also be used in retail stores and produce effective results.

References

[1] Softkenya. (2012). Kenya: SOFTKENYA. Retrieved from SOFTKENYA: softkenya.com/information/hardware-business-in-kenya/

[2] Clough, L. (2011). Marketing challenges and strategies for micro & small energy enterprises in East Africa. Global Village Energy Partnership International, 14-16.

[3] Pavlova, D. (2015). Customer equity management: The new business philosophy. *Trakia Journal of Sciences, 13*(1), 331-336.

[4] Kenneth, C. A., & Gabriel, J. M. (2013). Customer Relationship Management and Bank Performance in Nigeria: An Empirical Validation Study. *International Journal of Science and Research (IJSR)*, India Online ISSN: 2319-7064 , 416-422.

[5] Kracklaurer, A. H., & Wong .A. (2004). Customer Management as the Origin of Collaborative Customer Relationship Management. Journal of Collaborative Customer Relationship Management: Taking CRM to the Next Level, 3-6.

[6] Cheng, Y.I. Tang, K. Shen, R.J., & Hu, Y.H. (2005). Market Basis Analysis in a Multiple store Environment. Decision Support Systems, 339-354.

[7] Ramakrishnan, K. (2006). Customer retention: the key to business performance. Retrieved May, 2015 from: http://www.estrategicmarketing.com/smNov-Dec2/art1.1htm.

[8] Driskill, R., & Riedl, J. (1999). Recommender Systems for E-Commerce. Challenges and Opportunities, 1-4.

[9] Vozalis, E., & Margaritis, K. G. (2003). Analysis of recommender systems algorithms. In *The 6th Hellenic European Conference on Computer Mathematics & its Applications* (pp. 732-745)

[10] Hahsler, M. (2015). A Probabilistic Comparison of Commonly Used Interest Measures for Association Rules. Retrieved from: http://michael.hahsler.net/research/association_rules/measures.html

[11] Berisha-Shaqiri, A. (2008). The role of information technology in business growth performance in Kosova. *Challenges of the Knowledge Society*, 1-5