Predicting Factors which Determine Customer Transaction Pattern and Transaction Type Using Data Mining

Frehiywot Nega
HiLCoE, Computer Science Programme, Ethiopia
fr.nega@gmail.com

Tibebe Beshah
HiLCoE, Ethiopia
School of Information Science, Addis Ababa University, Ethiopia
tibebe.beshah@gmail.com

Abstract

The introduction of Mobile Phones dates back to 1995. Until recently, the use of mobile phones was mainly as a communication device. With the recent innovation, however, phone-based financial services are introduced as new features of mobile technology. Mobile money has been also introduced to Ethiopia in 2010. With the introduction of the service, Ethiopian banks and microfinance institutions are presented with new opportunities in which clients could make and receive payments using mobile devices.

The overall objective of this paper is to identify customer transaction pattern to improve the M-Birr system and to identify factors which determine the type of service used. In order to achieve the overall objective, the research employed CRISP data mining model, the practice of determining interesting knowledge from large amounts of data, which can be found from data warehouses, databases, or other information sources. The tools and application software used to accomplish the research are Association Rule Mining, Apriori algorithm, FP-Growth algorithm, Id3, J48, Prune, and WEKA. The review of the operation of M-Birr indicates that there are pre unexplored opportunities related to the use of the available transactions for improved business related decision making.

Keywords: Mobile Money; M-Birr; Association Rule Mining; Classification; CRISP

1. Introduction

Mobile money will allow Ethiopia to leap from the intermediate step between old paper-based systems and network of branches [5]. It brings financial services to the unbanked population of Ethiopia through branchless banking without incurring the substantial operating costs associated with the traditional banking approach [4]. M-Birr has been implemented over the past five years in Ethiopian financial sector, mainly in microfinance and very recently in one bank. Yet, there is not any system evaluation that has been done in terms of its adoption and use. As a result, much is not known about the extent and pattern of its adoption and use [2].

There is a huge electronic data stored in the financial transactions that could help an informed and systematic business related decision making. Because of the fact that, on the daily basis, there are large number of users as well as a large amount of financial transaction, the need for developing methods that analyze these transactions cannot be underestimated. The stored data should be manipulated using statistical techniques and software which are based on statistical theory. The processed data, coupled with the user’s judgment based on his/her experience and knowledge, increases the quality of decision making. To this end, some degree of simplification and automation is inevitable [1].

The review of the operation of M-Birr indicates that there are unexplored opportunities related to the use of available transactions for improved business related decision making like which M-Birr outlets are open on time, which M-Birr outlets close on time, which M-Birr outlets should be closed, where do we need more M-Birr outlets, which M-Birr outlets have a demographic profile and which are the peak times for transactions [3].

The M-Birr system has different functionalities which are deposit cash at an agent, deposit cash at branch, withdraw cash at an agent, withdraw cash at branch, transfer money, buy mobile top-up, pay bills, buy goods, repay loans, check balance, and get a statement. The M-Birr provider has indicated that the M-Birr system is not working as expected and the factors that determine the type of transaction remained unexplored. In order to make the system more accessible and reach the targeted potential customer, there is a need for empirical study that uncovers the transaction usage pattern and that identifies factors that determine the type of the services. A review of the system indicates that there are no expert analysts that could effectively use the available data for improving business related decisions and this study hopes to provide an expert analysis.

The available set of transactions are also large and yet not analyzed. Any attempt to do system evaluation using statistical methods poses a challenge [1]. The problem is the available transactional data are time-stamped and to be utilized it should be converted into time series data. Converting to time series data, however, is time consuming and not effective. Therefore, data mining techniques help to overcome the predicament of statistical methods by reducing the time it takes and being efficient. Besides, there are no regular updates about the status of the M-Birr operation. These problems should be turned into opportunities to enhance the value of the M-Birr operation.

The above-mentioned problems, however, are not insurmountable. Data mining techniques could help to effectively tackle the above-mentioned challenges [3]. Data mining is the process of searching and examining data in order to find hidden, but potentially useful, information [5]. This research employed the association rule and classification mining which are the most known data mining techniques [3]. Association rule is used to extract useful patterns from the M-Birr system customer transaction data, and classification is used to predict factors which determine type of transaction services [6].

2. Related Work

Electronic money (E-money) is a wider concept and refers to payments made using credit and debit cards, loyalty cards, prepaid cards, automated teller machine (ATM) cards, Near-field communication (NFC) enabled cards, store cards, gift cards and mobile phones [6]. Mobile money (M-money), thus, is a subsection of e-money. It refers to transactions and financial services completed on a mobile phone. A person who does not have personal bank account can use this service [2].

The term Mobile money (M-money) refers to the use of mobile phones to carry out financial and banking transactions. The scheme can be used to reach more than a billion people in the world who have little or no access to traditional financial services [1]. More than half households do not have access to financial services in the world [8]. The poor often must rely on informal financial services that may be more costly and less reliable. Low levels of financial presence present an obstacle to economic development. Availing mobile money services, however, ensures that users can receive funds securely, pay their bills, make bank transactions, transfer funds, and purchase goods and services. Consequently, in the development agenda financial presence has become an important topic [7].

Mobile Money facilities work by offering users an electronic wallet (e-wallet) service [6]. Access to the e-wallet is performed from the mobile phone and a number of basic functions can be performed such as putting money in e-wallet, withdrawing money from e-wallet, transferring money from one e-wallet to another, and checking balance or statement. Unlike saving deposits, e-wallets are interest free [10]. But users can opt to transfer their electronic money to a savings account whenever they want. Additional services will allow consumers to transfer money from their e-wallet into a loan account or even pay a utility bill or buy goods [9].

3. The Proposed Solution

3.1 Data Understanding and Collection

It is a bare fact that the concept of data mining doesn't exist without data. There is some real benefit if the data is already part of a data warehouse. If the data has already been cleaned for a data warehouse, then it most likely does not need further cleaning in order to be mined. Furthermore, in most of the data warehouses, many of the problems of data consolidation have already been addressed and maintenance procedures have been put in place. This does not mean, however, a data warehouse is a requirement for data mining. It is to emphasize the fact that setting up a large data warehouse that consolidates data from multiple sources resolves data integrity problems [3].

The data source for this research is the M-Birr system data of five independent service providers, each having its own M-Birr system data. The data stored is automated in M-Birr system. M-Birr system stores transaction data, customer information data, and agent information data. The total data obtained are 17,101 transactions from three different database tables, namely, agent registration table, customer registration table, and transaction table. The data has been stored since 2010. We selected age of customer, gender of customer, transaction year, opening time, closing time, time of transaction, type of transaction, transaction date, M-Birr outlet and amount of transaction and merged them into a new data table.

3.2 Feature Selection

After the initial data collection, the next step was describing the data set. The M-Birr data has 42 attributes of text, number, date and time formats. Among these attributes, 23 attributes are not accessible due to their private nature. The exclusion of these attributes, however, does not affect the result since they are not relevant for the realization of the envisioned research objectives. Nineteen attributes are taken from customer, agent and transaction data. Further refinement is done based on the research objective and in consultation with domain experts [7].

Attribute selection involves searching through all possible combinations of attributes in the data to find which subset of attributes works best for prediction. The best way to select relevant attributes is manually, based on a deep understanding of the learning problem and what the attributes actually mean as shown in Table 1 [1].

Table 1: Attribute Selection Table

No.	*Attribute Name*	*Data Type*	*Attribute Description*
1.	Opening time of the outlet	Time	This attribute describes the opening hour of the outlet. There are two kinds of outlets, namely, branch and agent outlet.
2.	Closing time of the outlet	Time	This attribute describes the opening hour of the outlet. There are two kinds of outlets, namely, branch and agent outlet.
3.	Time of transaction	Time	This attribute includes the transaction time described in the system as time using 3+ GMT.
4.	Gender of Customer	Character	This attribute includes the gender of the.
5.	Age of customer	Number	This attribute includes the age of the customer which is above 18 years old since in Ethiopia customers less than 18 are not allowed to open account.
6.	Type of transaction	Character	This attribute includes 11 different types of transactions.

No.	*Attribute Name*	*Data Type*	*Attribute Description*
7.	Transaction date	Date	This attribute includes the date of the transaction.
8.	Transaction year	Date	This attribute includes the year of the transaction.
9	MFI	Character	This attribute includes the MFI of the transaction.
10.	M-Birr outlet	Character	This attribute includes the outlet of the transaction.
11.	Amount of transaction	Number	This attribute includes the amount of transaction.

3.3 Data Cleaning

Since all the data in the M-Birr system are mandatory, the data is free from missing element. As a result, further data cleaning was not done.

3.4 Data Transformation for Analysis

Data transformation can involve smoothing or feature (attribute) construction. The overall intent is to minimize data redundancy. Smoothing techniques include binning, regression, and clustering. Attribute construction, on the other hand, is a process where new attributes are constructed and added from the given set of attributes to help the mining process. Smoothing can also serve as data reduction. For example, in the case of smoothing, the number of distinct values for a certain attribute is reduced [4].

Data transformation aims to manipulate the data so that its content and its format are suitable for the data mining process. In this paper, data transformation was done using smoothing through binning for time of transaction, type of transaction and transaction date.

4. Experimentation and Discussion

As presented in the previous section, the data is well understood, explored, selected and cleaned enough to be used for rule generation and model building. This Section presents the detailed activities carried out in selecting a modeling technique. It also identifies the implementation of the technique selected using the most appropriate algorithms in order to select the best rule for description and evaluation of the models and to find out the best one for prediction.

The study focuses on identifying determinant factors of customer transaction that lead to building a description model and prediction model. Association rule mining has been done using Apriori and FP-Growth algorithms. The rule is generated by changing the support and confidence value. Finally, the best rules are selected with acceptable confidence and support value. The classification models have been built using decision tree and PART rules. The models have been tested on the selected attributes and the significance of the outputs of the most important model is presented for analysis to domain experts. Finally, the model with the best performance is selected.

4.1 Experimental Setting

According to the CRISP data mining standard methodology employed in this research, selecting the actual association rule mining algorithm is the first step in modeling [2]. The most powerful associative and classification modeling are used. For association modeling, we included Apriori algorithm and FP-Growth algorithm.

Accordingly, Apriori and FP-Growth rule algorithm of WEKA were used to represent the knowledge/pattern identified. In an attempt to come up with significant rules, Apriori and FP-Growth algorithms were run on the M-Birr dataset with different support and confidence values. Different rules are generated using these algorithms by changing the support and confidence to discover the most interesting rules.

In order to accomplish this research part of classification, we used two data mining techniques. These are decision trees (using J48 and ID3 algorithm) and rule induction (using PART algorithm) for knowledge representation.

4.2 Result Evaluation and Comparison

Evaluation is a key step in any data mining process. It serves two purposes. The first purpose is the description of how well the final rule will work in the future and it is an integral part of many learning methods. The second purpose is to find the rule that best represents the data. In the series of experiments, evaluation of generated rule are selected in consultation with the domain expert and based on the soundness of the rules generated. The rules are:

Rule 1:

If Transaction_type = WC and Age = 24 and Gender_customer = Male

Then transaction date is the first ten days of the month

Rule 2:

If Transaction_type = TC AND Age = 40

Then transaction_date = third

Rule 3:

IF transaction date = third AND transaction type = DC

Then Amount_transaction = 300 Birr

Rule 4:

IF Amount_transaction = 300 Birr AND transaction type = WC

Then transaction_dat e= first

Rule 5:

If Amount_transaction = 300 Birr AND

Closing_time = 11PM AND

Time_transaction = first AND

Gender_customer = Male then WC

5. Discussion

The objective of this paper is to find interesting patter in the customer transaction of M-Birr system and predicting the factors that determine the type of transaction.

From the association rule mining, we found four interesting rules using Apriori algorithm on M-Birr customer transaction system data in terms of type of transaction, age, amount of transaction and gender of a customer. From the classification mining, we found four interesting rules by making type of transaction the target attribute with accuracy approximately 89.6 using J48 algorithm and rules with accuracy 89.51 using prune.

The research results indicate that there are practical ramifications for M-Birr in that it could further investigate the nature M-Birr system transaction. The M-Birr system has 11 different facilities. From the experiment we did, most commonly used ones are transferring money and withdrawing money but the main purpose of M-Birr is to be different from the traditional banking so more marketing campaign should be done in terms of the use of M-Birr. The M-BIRR system also revealed that, despite the fact that the demography of women and men is proportional, men dominate as customers. When M-Birr started in Ethiopia their target was to help people to use mobile phone and access their money easily but from the experiment we did the average payment is around 300 which indicates also that the pricing system should enable small payments.

6. Conclusion and Future Work

Managers generally do not have time to go through all records and data collections of their organization in order to make an informed decision. Moreover, the amount of data is bulky having several variables. In such contexts, it is extremely difficult to visualize patterns and relationships. As a result, managers of various levels need filtered and simplified data from their large amount of records. Knowledge discovery systems come handy to surmount the gap between the available huge data and limited time and thereby help managers of various levels to pass correct decisions or improve their plan. One tool is data mining that finds out hidden pattern from vast amount of data.

The research results indicate that there are practical ramifications for M-Birr operations. As the research indicated, despite the M-Birr system has 11 different facilities, the most commonly used ones are transferring money and withdrawing money out it. The implication is that the other facilities remain unutilized. Even these two frequent services, transferring money and withdrawing money, are conducted mainly within the first ten days and within the last 10 days. Uncovering the underlying reason requires an ethnographic approach so that M-Birr customers also remain active throughout the month. Based on the results, various campaigns that increase both the customer base and frequency by which the current customers are undertaking business via M-Birr should be launched.

Despite the demography of women and men is proportional, men dominate as customers. There has to be a further study that uncovers why men are the majority of the customers. Based on the result, a strategy should be designed to increase the customer base by appealing also to women. The fact that the average payment is around 300 indicates also that the pricing system should enable small payments. The M-Birr system, thus, should be designed in a way that it collects low-value deposits. The M-Birr system should revisit its pricing system to be conducive of small transactions. If the M-Birr system intent is to compete with traditional banks, which have branches, the system must show its user friendliness given that most of the people using this product are advanced in age and do not easily feel at ease with mobile technologies.

In this paper we did not look through why women are not users of the system so further study could be done to understand the reason why majority of the users are men.

References

[1] Aker, J. C., and Mbiti, I. M., "Mobile Phones and Economic Development in Africa". Journal of Economic Perspectives, 24(3), 2010, pp. 207-232.

[2] Merritt, C., "Mobile Money Transfer Services: the Next Phase in the Evolution of Person-to-Person Payments", Journal of Payments Strategy & Systems, 5(2), 2011, pp. 143-160.

[3] Donovan, K., "Mobile Money for Financial Inclusion", 2012.

[4] Asfaw, H. A., Financial Inclusion through Mobile Banking: Challenges and Prospects, 2015.

[5] Woldamariam, M. F., Ghinea, G., Atnafu, S., and Grønli, T. M., "Mobile Money System Design for Illiterate Users in Rural Ethiopia", In International Conference of Design, User Experience, and Usability, Springer International Publishing, 2014, pp. 482-491.

[6] Bramer, M., Principles of Data Mining, London. Springer, 2007.

[7] Mbiti, I., and Weil, D. N., "Mobile Banking: The Impact of M-Pesa in Kenya", National Bureau of Economic Research, 2011.

[8] Chen, M. S., Han, J., and Yu, P. S., "Data Mining: An Overview from a Database Perspective", IEEE Transactions on Knowledge and Data Engineering, 8(6), 1996, pp. 866-883.

[9] Fayyad, U. M., "Data Mining and Knowledge Discovery: Making Sense out of Data", IEEE Expert: Intelligent Systems and Their Applications, 11(5), 1996, pp. 20-25.

[10] Friedman, J. H., "Data Mining and Statistics: What's the Connection?", Computing Science and Statistics, 29(1), 1998, pp. 3-9.