Communication in Physical Sciences, 2020, 5(3): 291-298

An Improved Defragmentation Model for Distributed Customer's Bank Transactions

***C. K. Onwuamaeze and C. I. Ejiofor**

Received 17 April 2020/Accepted 20 May 2020/Published online: 24 June 2020

Abstract *The application of Information and Communication Technologies (ICTs) in trade and commerce has changed the ways trading transactions are carried out with the sole aim of accomplishing the growing expectations of organizational clients. In Nigeria nowadays, monetary transactions are made over disparate channels by bank clients. The exchanges can be heterogeneous, energetic, inter-related and as often as possible dispersed over numerous platforms. The variety, volume and velocity of the transactions can be very cumbersome for manual computations. In most cases, it is difficult to make informed decisions from the trend and patterns of the transactional datasets. However, a proper analysis on the datasets generated from the banking transactions of a customer can help in profiling the customer, target recommended solutions and achieve customer loyalty. This project implemented an improved defragmentation model for distributed customer's banks transactions that can be used by bank customers. It employed Naïve Bayes machine learning and collaborative filtering techniques to separate multiple transactions across numerous payment channels and deploy recommendations for the customer. The Prototype software methodology was adopted in the design. At the implementation of the research work, we utilized test cases to prove that customer's bank transactions over dispersed channels can be classified based on the user's query. Customers can now classify their transactions based on purpose of the transactions, the benefiting bank accounts, the beneficiary's name and their account numbers. The bank customer can quickly obtain a digital statement of account from all her bank accounts with the software.*

Key Words *Defragmentation, distributed bank transaction, financial transactions classifier, recommender System.*

***Chukwuemeka. K. Onwuamaeze**
Department of Computer Science, University of Port Harcourt, Choba, Rivers State, Nigeria
Email: *chukwuemekaonwuamaeze@gmail.com*
Orcid id:

Christopher. I. Ejiofor
Department of Mathematical Sciences, Edwin Clark University
Email: *ejioforifeanyi@yahoo.com*
Orcid id: 0000-0002-3466-8127

1.0 Introduction

As at today, there are twenty-one (21) commercial banks in Nigeria and nineteen (19) of them have individualized mobile banking software. The present banking software is designed to enable a customer operate a bank software tied to a bank account, make payment transfers from the account to other accounts, generate statement of accounts for the bank account it is tied to, and perform a few other utility functions.

Unfortunately, in a digitalized world, the present system shows that a bank customer has an account or accounts in one or several banks and must have different banking software to interface with the individual accounts. This scenario does pose several challenges to the bank users such that his computer system almost runs out of storage space due to the several applications duplicating functions. At other times, he may be confronted with the challenge of using different passwords for the different applications in his bid to escape the devastating effects of using the same password for multiple applications.

In today's banking software ecosystem, there is lack of interoperability among the several accounts of a single user - a clear absence of the principle of universality. A fundamental principle in software development is that a software designed in a domain should be able to perform similar functions in that domain. But in the case of Nigerian mobile software domain, no two software can be used together to accomplish any given task.

In a digital world, a person's record can be fetched from all its instances no matter how distributed or fragmented it may be. But that is not the here, there is no technology that aids the bank customer to aggregate all her bank transactions across different channels. And this negates the merits of computerization.

Furthermore, the present system does not have any defragmentation capability based on predefined criteria in such a way that he can obtain reports that show his financial status. In the present system, the user must approach each bank to obtain fragmented statement of accounts and then resort to manual defragmentation method. This method is archaic, anti-digital and retrogressive.

Lastly, the present system does not make any form of recommendation for the user. The user does not get any suggestion about available products and services from the banks. He is denied of the gain of knowing that there are cheaper, better and available alternatives to his previous preferences

2.0 Literature review

Multi-bank software based on MVC2 architecture and J2EE technology for customers with multiple bank accounts was designed by [1] to facilitate universality in banking operations. The system employed a three-level authentication mechanism namely, text-based, image-based and email-based authentication to ensure security in the banking operations.

Cardwatch developed by [2] featured neural networks trained with the past data of a particular customer to make the network process the current spending patterns to detect possible anomalies.

In the same line of thought, Neural Network and Classification Approach in Identifying Customer Behaviour in the Banking Sector: A Case Study of an International Bank developed by [3] proposed the use of Artificial Neural Network to new customer behaviour observations from previous customer observed behaviour after executing the process of learning from existing data. The study proposed a six-step procedure for classification of customer behaviours in the banking industry to predict which customer will stay (customer loyalty) with the bank and the ones that will change to other banks. Finie for (financial genie) designed and implemented by [4] is a voice-powered AI platform used to interact with a banking account using natural language queries. The application, which was based on deep neural networks is a remarkable natural language processing engine that has specifically trained its applications with a deeper knowledge of the financial and banking industry. Its machine learning capability enables the platform to expand knowledge and improve itself with every query. The study employed a combination of five AI engines such as LSTMs (long short-term memories) that enable the reasoning of the sequence of words using deep neural networks, the mapping of words to vector representations to understand the closeness of concepts, traditional recurrent neural networks and recurrent CRFs (conditional random fields) to extract meaning, among others.

A study conducted by [5] deployed Naive Bayes classifier for loan risk assessment. The authors worked on a database of 924 files of credits granted to industrial Tunisian companies by a commercial bank between 2003 -2006 The naive Bayesian classifier algorithm was used, and the results show that the good classification rate is of the order of 63.85 per cent.

Bank products and services recommendation can be integrated into the operational software of the bank. In a study carried out by [6], they created a web-based application that would grant expert advice on bank products to current and prospective customers. The specific bank products addressed include traveller's cheque, internet banking, smart kid savers, and mortgages. It will take queries from current and prospective clients, as the expert financial adviser would do, and mimic the expert in returning expert advice to the clients. The researchers deployed expert system methodology in carrying out the study.

In the same line of thought, a study on recommender system based on customer segmentation was carried out by [7] . The study extracted the recency, frequency and monetary variables of the clients and the variable's weights are calculated. Then using the weighted RFM and expectation maximization clustering algorithms and their combination with the closest-K-neighbours, recommendations for each cluster is independently extracted. The study found out that deployment of a hybridized algorithm in recommendation systems result in more accurate and precise predictions.

3.0 Materials and Methods

The methodology adopted for the development of this model is the prototyping method. The Prototyping Model is a systems development method (SDM) in which an archetype, an example of the final product is built, examined, and if necessary improved as necessary until satisfactory prototype is finally actualized from which the complete system or product can now be constructed. This model works best in situations where most of the project requirements are not known in detail at the commencement of the project. It is a repetitive, trial-and-error process that takes place between the developers and the users. The choice of this method is hinged on the fact that, first of all, all the requirements of the project were not known at the start of the project and secondly, there was no available live account for demonstration purposes.

The proposed system has been designed to ameliorate the challenges of the present system. Having looked at the weaknesses of the present, system, we developed the proposed system to overcome these challenges and weaknesses and present the bank customer with a robust financial transactions classifier.

The software works as a multi-bank accounts aggregator, interfacing between the user and her multiple bank accounts in one or several banks. This enables a bank user to use the software to perform the universal function of one-to-many solution.

The present system lacks the capacity to perform defragmentation functions to enable the customer generate decision-making information, but the proposed system will serve as a defragmentation utility program that will enable bank customers to query their accounts, filter and collate results from transactions distributed across multiple channels. Customers can now classify their transactions based on purpose of the transactions, the benefiting bank accounts, the beneficiary's name and their account numbers. The bank customer can quickly obtain a digital statement of account from all her bank accounts with the software.

The proposed system will also serve as recommendation software that will provide product and service recommendations to customers based on similarity proximity between the profile of the customer and the profile of their transactions. As a recommendation system, it will help the customer to know about available, cheaper, more qualitative alternatives to their previous preferences.

Table 1 presents the list of Input and Output documents used in developing the software application.

3.0 Input and Output Document

Table 1: Input and Output format table

S/N	Input	Output
1	Purpose	Defragmentation by Purpose of Transaction
2	Beneficiary	Classification by Payee's Name
3	Deposit	Credit Balance
4	Use Frequency	User's Category
5	Recipient's Account Number	Classification by Payee's Account Number
6	Payer's Account Number	Classification by Payer's Account Number
7	Banks	Classification by Payee's Bank
8		Statement of Accounts
9		Recommendations

They specify what a user can input into the software and the resultant output information that can be expected. Furthermore, this software will enhance the aggregation of the customer's financial records irrespective of how distributed or fragmented they may be. This software will bring to an end a situation where people use different and fictitious names to commit financial fraud without having any technology to link the records.

Finally, the application will self-monitor the customer's transactions to observe when there are unusual activities on the accounts. This will be achieved using triggers on the program, which, when the threshold is crossed, the user is alerted on the unusual activity being noticed in the account(s).

3.1 Defragmentation functions

The user can defragment the transactions based on the following parameters: Account Name, Account Number of the paying accounts or the receiving accounts, Bank Names and the Purpose of the transaction. Thus, the user can decide to query the system to filter all transactions made to a particular individual and the system will generate a report showing all transactions made in that name across the multiple channels.

3.2 Algorithm for defragmentation by Payee

Step 1: The user makes transactions
Step 2: The system learns the transactions
Step 3: The system classifies the transactions
Step 4: Get the total amount of transactions
Step 5: The system classifies the names of the payee
Step 6: The system matches the total transactions to the classified names
Step 7: Display result

3.3 Algorithm for defragmentation by account numbers

Furthermore, the user can query the system to defragment the transactions based on the bank account numbers of the receiving bank using these machine learning techniques:

4 Step 1: The user makes transactions
5 Step 2: The system learns the transactions
6 Step 3: The system classifies the transactions
7 Step 4: Get the total amount of transactions
8 Step 5: The system classifies the account numbers of the beneficiaries in the database
9 Step 6: The system matches the total transactions to the classified account number
10 Step 7: Display result

3.4 The Use of Naïve Bayes as a Classifier in Defragmentation of Customer Transactions

The system is able to perform defragmentation after it has learned the spending patterns of the user which is stored in the database. The learning is done using Naïve Bayes classification model. This model makes an independent assumption which means that knowing the value of one attribute does not tell the value of another attribute between predictors.

The Bayesian equation is of the form

$$P(c/x) = \frac{P(x/c)\ P(c)}{P(x)} \quad (1)$$

This equation can be interpreted to mean that the probability of the class given the data (predictor) is equal to the probability of the data given the class multiplied by the probability of the class divided by the probability of the data (predictor).

where **P(c/x)** = The posterior probability of the class given the data (predictor//attribute)

P(c) = The prior probability of the class

P(x/c) =The Likelihood of the predictor given the class

P(x) = the prior probability of the data (predictor/attribute)

The Application of Naïve Bayes Algorithm

The NB algorithm can be spelt out as

Step 1: Construct a frequency table for each item (attribute) and the group.

Step 2: Then transform the frequency tables to likelihood tables.

Step 3: Use Naïve Bayes equation (1) to compute the posterior likelihood of each group

Step 4: The class with the highest posterior probability is the outcome of prediction.

3.5 Application of content-based filtering algorithm in recommender systems

The proposed system also employs content-based filtering algorithm to recommend products and services to the user. The system learns the spending patterns of the customer using the Naïve Bayes classifier and then analyses the history of the customer's transactions. From that study, it compares similarity of the user profile to the similarity of the products to make recommendations. The recommender system generates product and services recommendations from a large collection of items based on the bank user's previous transactions. The Content-based technique emphasizes more on the analysis of the attributes of items in order to generate predictions.

In this project, we concentrate on applying a suitable measure of lookalike within products and user's profile. The algorithm deployed seeks to find products and services that are closer to a particular user profile. There are numerous similarity-measurement techniques that can be applied, but we choose to use the cosine similarity, Fig. 1 shows a unified synopsis of the interplay of different components that make up the Improved Customer's Bank Transactions Defragmenter.

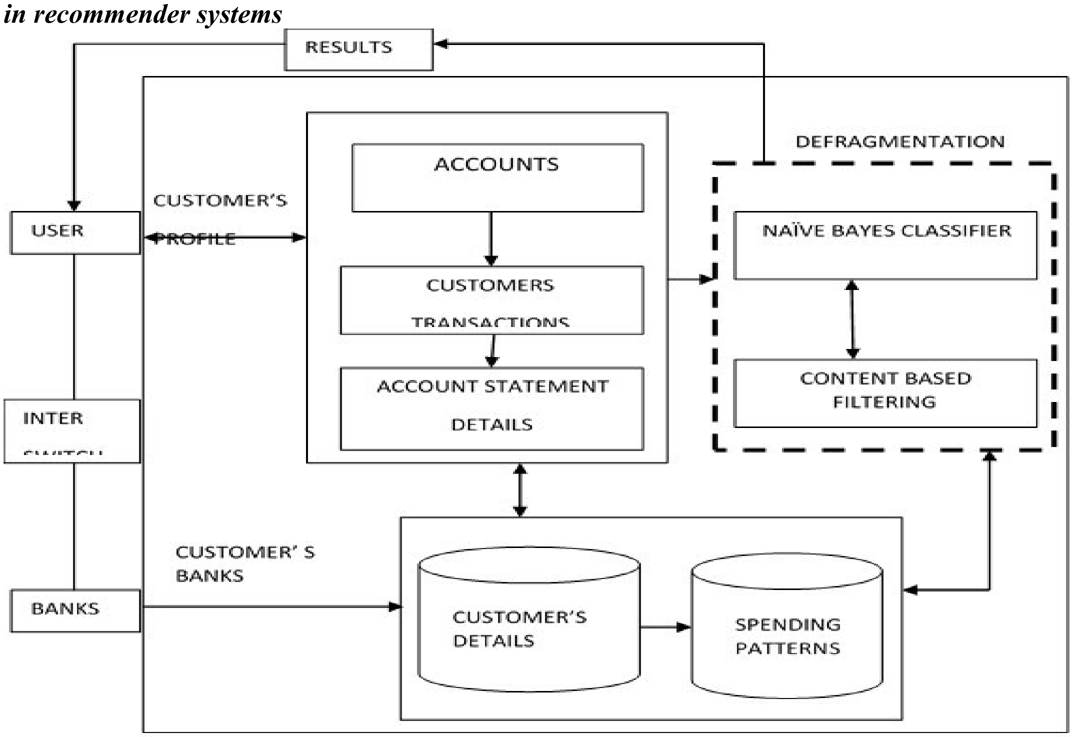

Fig. 1: An Architecture of Defragmentation Model for Distributed Customer's Bank Transactions

It considers items as document vectors of an n-dimensional space and computes their similarity as the cosine of the angle that they form.

Precisely, we have considered cosine-based similarity as a measure of distance between a user profile and an item profile. Hence if pm = (pm1, …, pm, L) is the profile of user u_m and q_n = (qn, 1, …, qn, L) becomes the identifier of the product using the identifiable similarity between the two identifiers

$$\mathbf{S(p_m, q_n)} = \frac{\sum 1pm,1,qn,1}{\sqrt{\sum 1p^2m,l},\sqrt{\sum_1 q^2n,l}} \qquad \mathbf{(2)}$$

Steps in Content Based Filtering Algorithm

Step 1: Set User Profiles (UP_1, UP_2, UP_3 … UP_n)

Step 2: Set Products Profiles (PP_1, PP_2, PP_3 … PP_n)

For every Product Profile (PP_1, PP_2, PP_3 … PP_n) similar to User Profile (UP_1, UP_2, UP_3 … UP_n) make a match

Step 3: Record lookalikes in matrix UP x PP for all UP's (M_1, M_2, M_3, … M_n)

Step 4: Record total lookalikes

Step 5: Match all UP's with their corresponding PP's

Step 6: Display Recommendation

IV) COMPONENTS OF THE ARCHITECTURE

The diagram above illustrates the architecture of enhanced bank software showing the proposed defragmentation system in dotted line. The system shows the user and his interaction with the software and the banks.

THE CUSTOMER'S BANKS MODULE

The banks can register with the system, provide and view own profile, view customer profile, view spending patterns of customers and deploy recommendation of products and services to target customers.

ACCOUNTS MODULE

The accounts module enables the bank customer to register as many bank accounts as possible. The user specifies the bank name, the account name and number, the starting balance on the account, the maximum deposit and maximum withdrawal from the account.

The Customer's Transactions Module

This is the module that allows the user to make payments for goods and services to other customers who have bank accounts. To make a transaction, the user selects the purpose of the transaction, the beneficiary's account details and the bank account(s) from which the payment is to be made. The user also enters the security code (an internal check) to ensure that only genuine and authorized transactions are processed.

The Accounts Statement Details Module

This module allows the user to generate bank account statements from different banks as specified by the user.

The Interswitch Module

The Interswitch serves as the bridge between the bank customer and the bank services. It facilitates e-payment and e-commerce. It enables fund transfers from one bank account to another.

The Defragmentation and Learning Module

This is the module that enables defragmentation functions in the proposed system. The user can defragment the transactions based on the following parameters: Account Name, Account Number of the paying accounts or the receiving accounts, Bank Names and the purpose of the transaction.

The Naive Bayes Classifier Module

This module is responsible for the classification operations of the system. The program is able to classify the customer's bank transactions with the intelligent machine residing in this module.

Table 2 is a snapshot of unfragmented Account Statement of a Bank customer. It is generated from different bank accounts of the customer. The transactions were made over different periods of time and were made for different purposes and to different beneficiaries. However, Table 3 shows a sample of defragmented Account Statement of a Bank Customer using the Bank Transactions Defragmenter.

Table 2: The unfragmented Account Statement Sample

Tranx date	Acct name	Acct id	Product id	Product name	Amount
01/03/2018	FBN - 0012	1223456789	1	Clippers	2000
01/03/2018	FBN –0012	1223456789	2	Tithes	3000
01/03/2018	FBN –0012	1223456789	3	Men Shoes	10000
01/03/2018	FBN –0012	1223456789	4	Property (Land, House)	300000
01/03/2018	FBN –0012	1223456789	2	Tithes	2000
01/03/2018	FBN –0012	1223456789	5	Phones	15000
01/03/2018	FBN –0012	1223456789	6	Men Suit	10000

01/03/2018	FBN –0012	1223456789	7	Children Shoes	5000
01/03/2018	FBN –0012	1223456789	8	Computers	70,000
01/03/2018	FBN –0012	1223456789	1	Clippers	2000
01/03/2018	FBN –0012	1223456789	9	Rent	200000
04/04/2018	FBN –0012	1223456789	10	Transfer to Brother	10000
04/04/2018	FBN –0012	1223456789	10	Transfer to Sister	2000
04/04/2018	FBN –0012	1223456789	11	Meat	2000
04/04/2018	FBN –0012	1223456789	13	Chocolate Tea	550
04/04/2018	FBN –0012	1223456789	14	Roll of Milk	250
04/04/2018	FBN –0012	1223456789	15	Rice	1600
04/04/2018	FBN –0012	1223456789	16	Beans	1000
04/04/2018	FBN –0012	1223456789	17	Upkeep Prince	5000
04/04/2018	FBN –0012	1223456789	18	Plagiarism Test	5000
01/03/2018	FDL-002	1234567890	1	Clippers	500
01/03/2018	FDL-002	1234567890	2	Tithes	1000
01/03/2018	FDL-002	1234567890	2	Tithes	2000
01/03/2018	FDL-002	1234567890	3	Men Shoes	5000
01/03/2018	FDL-002	1234567890	4	Property (Land, House)	200000
01/03/2018	FDL-002	1234567890	2	Tithes	2000
01/03/2018	FDL-002	1234567890	5	Phones	20000
01/03/2018	FDL-002	1234567890	7	Children Shoes	6000
01/03/2018	FDL-002	1234567890	8	Computers	30,000
01/03/2018	FDL-002	1234567890	1	Clippers	4000
01/03/2018	FDL-002	1234567890	9	Rent	100000

Table 3: The Defragmented Account Statement Sample

Property (Land and House)

TRANX DATE	ACCT NAME	ACCT ID	PRODUCT ID	PRODUCT NAME	AMOUNT
01/03/2018	FBN –0012	1.223E+09	4	Property (Land, House)	300000
01/03/2018	FDL-002	1.235E+09	4	Property (Land, House)	200000
14/05/2018	FBN –0012	1.223E+09	4	Property (Land, House)	55000

17/05/2018	FDL-002	1.235E+09	4	Property (Land, House)	228000

School Fees

Tranx date	Acct name	Acct id	Product id	Product name	Amount
13/05/2018	FBN –0012	1.223E+09	21	School Fees - Prince (UNN)	70,000
13/05/2018	FBN –0012	1.223E+09	22	School Fees - Favour (FUTO)	46000
14/05/2018	FBN –0012	1.223E+09	26	School Fees - Miracle	25000
14/05/2018	FBN –0012	1.223E+09	27	School Fees - Success	23500
14/05/2018	FBN –0012	1.223E+09	28	School Fees - Chuks (UNIPORT)	186,200

Groceries

Tranx date	Acct name	Acct id	Product id	Product name	Amount
04/04/2018	FBN –0012	1.223E+09	11	Meat	2000
04/04/2018	FBN –0012	1.223E+09	13	Chocolate Tea	550
04/04/2018	FBN –0012	1.223E+09	14	Roll of Milk	250
04/04/2018	FBN –0012	1.223E+09	15	Rice	1600
04/04/2018	FBN –0012	1.223E+09	16	Beans	1000
03/05/2018	FBN - 0012	1.223E+09	15	Rice	22000
04/05/2018	FBN –0012	1.223E+09	16	Beans	3000
04/05/2018	FBN –0012	1.223E+09	11	Meat	1500
10/05/2018	FBN –0012	1.223E+09	14	Rolls of milk	1000
14/05/2018	FBN –0012	1.223E+09	11	Meat	2000

Upkeeps

Tranx date	Acct name	Acct id	Product id	Product name	Amount
04/04/2018	FBN –0012	1.223E+09	10	Upkeep for Brother	10000
04/04/2018	FBN –0012	1.223E+09	10	Upkeep for Sister	2000
04/04/2018	FBN –0012	1.223E+09	17	Upkeep Prince	5000
12/05/2018	FBN –0012	1.223E+09	17	Upkeep Prince	10000
13/05/2018	FBN –0012	1.223E+09	20	Upkeep Favour	10000
14/05/2018	FBN –0012	1.223E+09	29	Upkeep Wife’s	30000
15/05/2018	FDL-002	1.235E+09	30	Upkeep Mother	10000
15/05/2018	FDL-002	1.235E+09	31	Upkeep Mother-in-law	10000
16/05/2018	FDL-002	1.235E+09	17	Upkeep Prince	10000
16/05/2018	FDL-002	1.235E+09	20	Upkeep Favour	10000
26/05/2018	FDL-002	1.235E+09	34	Upkeep Personal	15000

The Recommender System Module

The recommendation function of the system resides in this module. The system learns the spending patterns of the customer using the Naïve Bayes classifier and then analyses the history of the customer's transactions. From that study, it compares similarity of the user profile to the similarity of the products to make recommendations.

The Customer's Details Module

This module holds the database of the details of the bank user. When the customer queries the system, he would be prompted for his password. If the password is validated, the customer can access his banks details and make transactions. The banks can also profile a customer showing personal details, maximum deposit in the account, maximum withdrawal limit and the internal security code that self-monitors the activities in the account.

The Spending Patterns Module

This module integrates machine learning algorithm to the defragmentation system. It is the part of the software that makes the system intelligent and performs autonomous functions. The system studies the spending patterns of the user using Naïve Bayes classifier system. It can, therefore, classify a user into categories of a low, medium or heavy user. The system is trained to recognize variations in the spending patterns of the user before making the classification.

The Results Module

The customer is able to fetch result to queries made to the system through this module. The results can be in form of bank statements, spending patterns, payment beneficiaries, etc.

4.0 Conclusion

Prior to the development of this software, no bank software in Nigeria was able to assist the customer in filtering bank transactions made across multiple payment channels, but the proposed system has pioneered classification solution to heterogeneous, dynamic, inter-related and distributed bank transactions no matter their variety, volume and velocity. This software can filter out transactions based on four parameters, namely, the purpose of the transaction, the name of the payment beneficiary, the bank account number of the beneficiary plus brand name of the beneficiary's bank. Indeed, **it** is a handy tool for mining, capturing, filtering and extracting distributed financial records.

5.0 References

Agrawal, V., Kumar, V, Ghodke, V. & Temdhare, K. (2014) Multibanking Transaction Interface using MVC2 Architecture. International *Journal of Computer Science & Engineering Technology (IJCSET)*, 5, 4, pp. 284-288

Aleskerov, E.,, Freisleben, B. & Rao, B. (1997) CARDWATCH: *A Neural Network Based Database Mining System for Credit Card Fraud Detection*. In: Proceedings of Computational Intelligence for Financial Engineering (CIFEr), 220--226,

Ogwueleka, F. N., Sanjay, M., Ricardo, C. P. &, Luis, F. Neural Network and Classification Approach in Identifying Customer Behaviour in the Banking Sector: A Case Study of an International Bank, *Human Factors and Ergonomics in Manufacturing and Setvice Industries,* 25, 1, pp. 28-42.

Amit C, howdhry (2016) How Clinc Developed The Most Advanced A.I. for the banking Industry available at https://www.forbes.com/sites/ -amitchowdhry/

Krichene, A. (2017). Using a naive Bayesian classifier methodology for loan risk assessment: Evidence from a Tunisian commercial bank., *Journal of Economics, Finance and Administrative Science*, 22, 42, pp. 3-24.

Adebayo, A, O., Agbola, I . S., Ayangbade, A. O. & Obajimi, O. O. (2015). Bank Products Recommender. *The International Journal of Engineering and Science*, 4, 7, pp. 41-52.

Raamani, R. (2016). Recommender system based on customer segmentation (RSCS). *Kybernetes*, 45, 6, pp. 946-961.