Survey and Challenges of Mining Customer Data Sets to Enhance Customer Relationship Management

Raghda M. Nasr
CIS Dep., Sadat Academy for
Management Sciences

Sahar Nagaty
BA Dep., Sadat Academy for
Management Sciences

Mohammed B. Senousy
CIS Dep., Sadat Academy for
Management Sciences

ABSTRACT
Data mining is a useful and powerful tool for any organization especially for marketing people. Data mining is used in managing relationships with customers. New trends in customer relationship management –CRM- have growth. One of them is social customer relationship management –sCRM-. The major problem here is data accessibility. If the research process is not started from inside the organization, the researchers cannot have access to customer's data. In this research paper, the research creates a survey to find out available customer's data sets that can be used in academic research. In addition, the research reviews basic concepts of CRM and Data Mining Techniques. Finally, the research focuses on current challenges of mining customer's data from business perspective and analytical customer relationship management -aCRM-.

General Terms
Data mining, Customer relationship management

Keywords
CRM, aCRM, Data Mining, sCRM, i-CRM

1. INTRODUCTION
Most of businesses agreed that customers are the most valuable resource which guarantees the businesses continuing success. Therefore, businesses pay effort to manage its relations with its customers. This process called Customer Relationship Management "CRM". CRM can be classified into three categories Operational CRM, Collaborative CRM and Analytical CRM [1]. Operational CRM improve the efficiency of the business processes involving customer services, order management, sales and marketing automation and management. Collaborative CRM involving all communications, coordination and collaboration processes with the customers. Analytical CRM "aCRM" improve the company's relationship with the customer, by involving the activities that capture, store, manage, evaluate and report customer's data.

Analytical CRM is "Applying business analytics techniques and business intelligence such as data mining and online analytic processing to CRM applications" [2]. In other words analytical CRM is the interpretation of data collected by the operational CRM to identify the opportunities, optimize customer interaction, and manage business performance. Analytical CRM takes the data gathered from sources such as marketing campaigns, and products group then runs algorithms over it for analysis and interpretation purposes. Analytical CRM has been used to optimize profitability, revenue and customer satisfaction analysis, customer profiling and categorization, up/cross selling, fraud analysis and churn management. One of the most used tools for aCRM is data mining.

Data Mining "DM" is the process of extracting knowledge and discovering new patterns and relationship by digging into

large amount of data [3]. DM techniques may help CRM in its goal by extracting and finding the customers characteristics. In other words the data mining may be helpful in determine the behaviors of the customer during specific stage in Customer Life Cycle or determine the other customers with the similar behavior pattern

This research organized as follows, Section 2 reviews CRM concepts and related researches. Section 3 reviews data mining techniques in CRM. Section 4 states current challenges in mining customer's data sets. Finally, conclusions are given in Section 5.

2. CRM RELATED RESEARCHES
CRM was defined according to Peelen as "A process that addresses all aspects of identifying customers, creating customers knowledge, building customer relationships and shaping their perceptions of the organization and its products" [1]. Where Turban defined CRM as "A Customer Service approach that focuses on building long-term and sustainable customer relationships that add value both for the customer and the selling company" [2] . Another definition by Adela Tudor "CRM is a bucket of IT Applications and procedures whose target is to identify the main expectations and preferences of the clients and to use efficiently the gathered information in order to improve the relationships between the business and the customers" [4]. According to Farooqi CRM is "The Process of Acquiring, relating and growing profitable customer which requires a clear focus on service attributes that represent value to the customer and create loyalty" [5]. From the previous definitions the researcher concluded that CRM is the business strategy which aims to develop a profitable relationship between customers and the organization in order to gain a competitive advantage to achieve success.

Any CRM process has four components and four dimensions. CRM components are stated to provide a clear guidance during the relationship with customers. Any organization should pay time and effort to collect data about customers, determine the communication channels they prefer, then prepare a relationship strategy and finally know when their customers need a customized product and provide it. From another point of view the CRM dimensions can be considered as a framework for the customer life cycle within the organization.

2.1 CRM Dimensions
According to [4] and [6], CRM consists of four dimensions:

2.1.1 Customer Identification:
It is also called Customer acquisition. The aim of this phase is to know who are the most likely to be a valuable customer and target him. Target customer analysis seeks the profitable customer through the analysis of customers' underlying characteristics, where the segmentation is the subdivision of

the customer database into small groups of customers who are relatively similar.

2.1.2 Customer Attraction:
After identifying the target customer, definitely the organization will focus all the effort and resources to attract the profitable customer. This will be by using the direct marketing, which motivate the customer through different channels (ex: direct mail, sms, coupons, etc...).

2.1.3 Customer Retention:
The best way to keep the customers, and restore customers is to increase their satisfaction. Customer retention is considered as the CRM core. Customer retention elements are one to one marketing, loyalty programs and complaints management.

2.1.4 Customer Development:
This phase aims to increase the size of the customer transactions with the company. Elements of customer development include customer lifetime value analysis, up/cross selling and market basket analysis. The customer lifetime value is the total revenue expected to be gained from the customer during his relationship with the organization. Up/Cross selling are promotions activities aim to market related services according to the customer use. Market basket analysis maximizes customer transactions by revealing regulations in the customers' purchase behavior.

2.2 CRM Components
CRM consists of four elements of a simple framework: Know, Target, sell and service [7]. The pervious definition can summarize the CRM components to the basic case, but it will neglect a lot of basics that should be well known for any organization cares about its customer relationship management. Figure 1 illustrates these components in detailed levels.

First, organizations should gather enough information about its customers. Knowledge is the currency of the interactions between the organization and its customer, when efficiently exploited it can become a competitive advantage [8]. All the knowledge about the customers is stored in his profile (Customer Profile). Most of organizations stopped at this point, the collected information about the customer is enough to be collected during his life cycle within the organization. Customer Knowledge passes by three profiles: customer profile, user profile then human profile. The basic information fill in the customer profile such as his name, address, phone no, segment …etc. The user profile includes more information about his way of using the product or service, his complaints, transactions. By the end any organization should prepare to extend the profile into human profile. Some organizations stored his preference for a particular communications channel; others extend this part till his preferred color, hobbies or even entertainment destinations. Any organization should know how to use this knowledge to make customer feel as family member in the organization.

Finally, it is clear that the relationship strategy is not a sales strategy. Sales strategy end at the moment the transaction is completed. But relationship strategy has a long term interest in customers. Relationship strategy may be stated according to the customers segments. Another aspect may be according to the relationship phase. During customer life cycle there are four relationship phases: exploratory, growth, maturity and decline, relationship strategies for one differ from the other. Any Company should be able to identify the target customers segment and their relationship phase, so it will be able to choose the suitable relationship strategy. Company should provide loyalty programs to keep the customer in maturity phase as long as possible. Customer loyalty was measured by purchasing pattern [9], while it should not be an indicator as the customer may purchase more from other competitors. Loyalty should be measured by more than one indicator such as share of wallet, frequency of visits, vital range of products, all of the previous with one of the most indicators his satisfaction.

2.3 New Trends in CRM
The beginning of CRM was only one direction from the business to customer. By the time the CRM evolved to a bidirectional communication. Today the relationship between customers and business is influenced by others customers with their word of mouth also all the organization communications will be exposed to the other competitors. CRM now is a multidirectional communications. Facebook, Twitter, and all the other social media opened a numerous channels for all the parties to contact at any time. The future of CRM will be social transparent and customer centric. SCRM (Social Customer Relationship Management) in these coming days is about customer engagement, not customer management [10]. Also we can state that SCRM is about including customers in the activities of the organization [11].

SCRM is a process to monitor, engage and manage conversation and relationships with existing and prospective customers and influencers across the internet, social networks and digital channels [12]. Organizations may use SCRM to communicate with customers, respond to customers question, promote events, capture customer data or provide support [13]. Customers are connecting and having conversations about enterprises through social media, whether enterprises are actively engaged or not, and brands risk being raised if they fail to participate [12].

Intelligent Customer relationship management "i-CRM" is another direction. i-CRM has been evolved a decade ago. Baracskai & others, construct an intelligent portal that can support managing Customer relationships.

This intelligent portal helps customers to get the requested information quickly and customizes questionnaires for each customer [14]. Another research was about supporting customer involvement in product and service development by i-CRM [15]. From another perspective, [16] presents a new conceptual framework and practical solution for e-loyalty by using i-CRM on the cloud, by trying to minimize negative feedback of a customer and increase e-loyalty of an organization by enhancing the CRM interactivity. i-CRM could be concluded as a mix of business intelligence and machine learning to support decision makers [17].
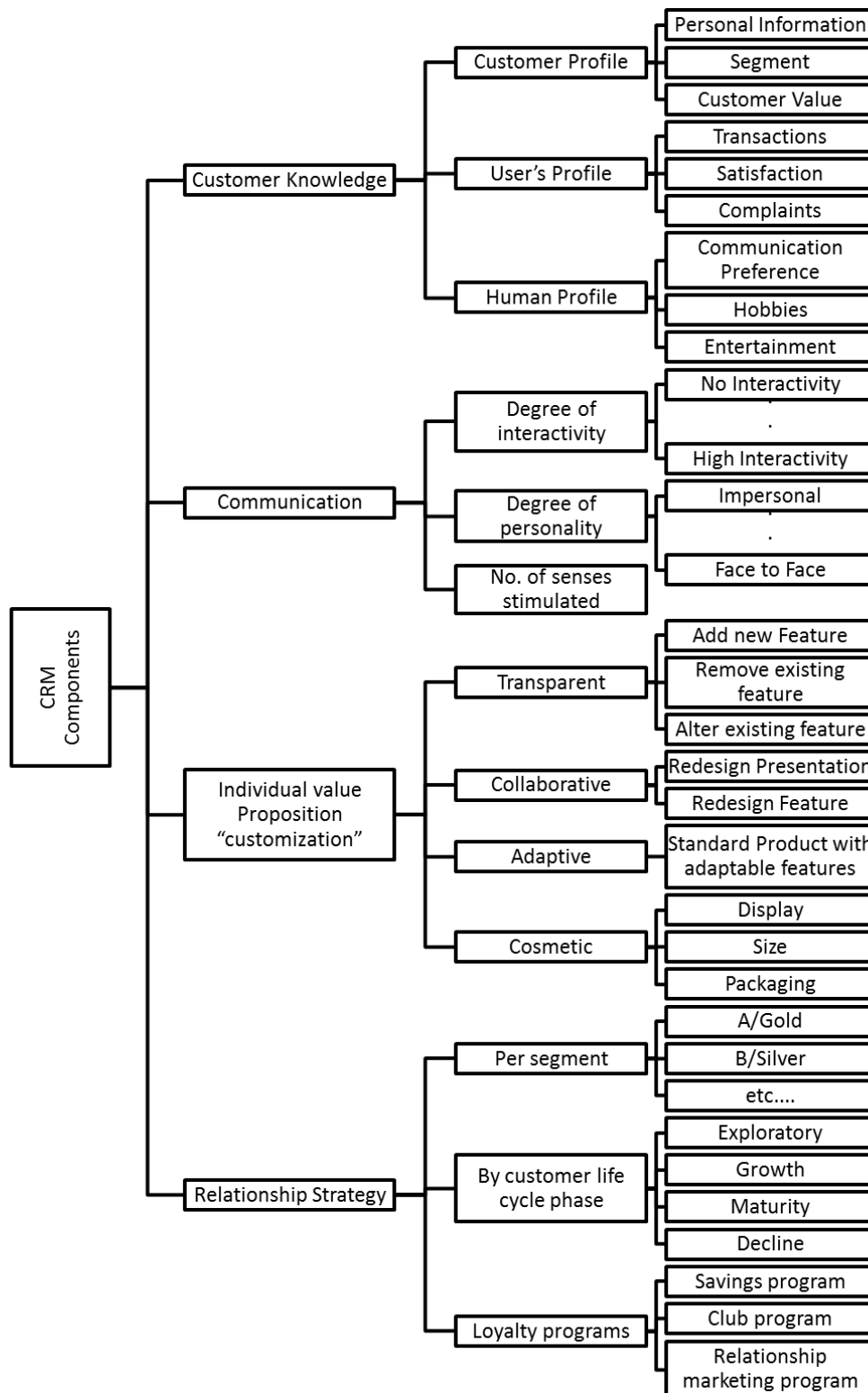
Figure 1: CRM Components
Source: Illustrated by the Researcher

3. DATA MINING IN CRM

"Data mining is a business process for exploring large amounts of data to discover meaningful patterns and rules" [18]. Data mining is an interdisciplinary field that includes database systems, statistics, machine learning, visualization, and information science [3]. Data mining techniques deal with discovery and learning [5]. Data mining may be helpful to accomplish the goal of CRM by extracting or detecting hidden customer characteristics and behaviors from large databases.

3.1 Data Mining Techniques and Tools

There are a lot of data mining techniques. These techniques could be predictive or descriptive. Predictive techniques include classification, Regression analysis. Descriptive Techniques includes clustering, association rules, sequential pattern discovery and visualization. All these techniques could be used in aCRM. **Error! Reference source not found.** describes each technique.

Table 1: Data mining techniques

Technique	Type	Description
Classification	Predictive	Technique used to predict group membership for data instances. Popular classification techniques include decision trees and neural networks
Regression analysis	Predictive	It is a statistical technique that studies the linear relationship between variables usually under an assumption of normally distributed errors.
Clustering	Descriptive	Technique used to place similar data records into related groups.
Association rule discovery	Descriptive	Technique created by analyzing data for if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true.
Sequential pattern discovery	Descriptive	Technique in which each object associated with its own timeline of events; find rules that predict strong sequential dependencies among different events.
Visualization	Descriptive	The interpretation of relationships in multidimensional data. Graphics tools are used to illustrate data relationships

3.1.1 Data Mining Tools Survey

Table 2: Data Mining Tools

Tool Name	Author / Group / Foundation	Written in	URL	Source	Released In
Weka	Machine Learning Group- University of Waikato	Java	http://www.cs.waikato.ac.nz/ml/weka/downloading.html	Open	2002
	Weka support data pre-processing, classification, regression, clustering, association rules, and visualization.				
Orange	Faculty of Computer and Information Science, University of Ljubljana	Python	http://orange.biolab.si/download/	Open	2013
	Orange support preprocessing, Classification, regression, Evaluation, Association, unsupervised techniques, visualization and visual programming.				
RapidMiner studio / RapidMiner sever	RapidMiner, Inc.		http://rapidminer.com/pricing/ Require an account for download	Closed	
	SVM , Decision Trees , Rule Learner , Lazy Lerner, Bayesian Learner , logistics learner , Association rules , and Clustering.				
R	Robert Gentleman and Ross Ihaka, Statistics Department of the University of Auckland Developed at Bell Laboratories	It is a language itself	http://cran.r-project.org/mirrors.html	Open	1997
	Association Rules & Frequent Itemsets , Sequential Patterns , Classification & Prediction, Regression , Clustering , Time Series Analysis , Text Mining , Outlier Detection , Social Network Analysis and Graph Mining , Spatial Data Analysis , Statistics , Data Manipulation and Data Access				
Apache Mahout	Apache Foundation	Java	http://mahout.apache.org/general/downloads.html	Open	2010
	Classification, Clustering and recommendations algorithms for distributed machine learning				

Error! Reference source not found. shows a survey of most well-known data mining tools including author, source type, release date and supported techniques in each tool.

3.2 Data Mining Techniques for CRM Dimensions

Each one of CRM dimensions can be divided into one or more task. Customer identification includes target customer analysis then customer segmentation, then Customer Attraction is completely focused in one task which is the direct marketing. Customer Retention divided into three tasks: Loyalty programs, 1:1 marketing and Complaints Management. Finally, Customer development consists of customer lifetime value, up/cross selling and market basket analysis. As mentioned in section 3.2 data mining techniques varying from each other. The tasks of CRM dimensions may be achieved by more than one technique; but there are some techniques works well than the others. Table 3 shows the techniques that can work better for each task. The table has been extracted from several previous researches such as: [4], [5], [6] and [7].

Table 3: Data Mining Techniques for CRM Dimensions

CRM DIMs / DM Techniques	Customer Identification		Customer Attraction	Customer Retention			Customer Development		
	Target Customer Analysis	Customer Segmentation	Direct Marketing	Loyalty Programs	1:1 marketing	Complaints Management	Customer Life Time Value	Up/cross Selling	Market Basket Analysis
Classification	✓	✓	✓	✓	✓		✓		
Clustering	✓	✓	✓	✓	✓	✓	✓		
Forecasting							✓		
Regression		✓	✓	✓			✓		
Sequence Discovery				✓		✓		✓	✓
Visualization	✓								
Association					✓			✓	✓

3.3 Data Sets Survey

There are a lot of sites that present the descriptions for a lot of data sets about customers in many fields but usually the data is found offsite or unpublished, as the customers profiles considered as a very sensitive data. Most of the sites that provide the data set make it available only for the purpose of study, so they must get confirm that the person who submit the request is registered in any University. The following table shows 6 data sets about customers, domain and availability.

Table 4: Data Sets Survey Table

Data Set Owner	URL	Data Domain	Data Description	Available
Machine Learning Repository	https://archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data	Restaurant & consumer data Set	The dataset was obtained from a recommender system prototype. The task was to generate a top-n list of restaurants according to the consumer preferences.	Free
AllState	https://www.kaggle.com/c/allstate-purchase-prediction-challenge/data	Allstate Purchase Prediction Challenge Insurance	The data set shows the transaction history for customers that ended up purchasing a policy. For each customer_ID, there are their quote histories. In the training set the entire quote history are available, the last row of which contains the coverage options they purchased. The test set, provide only a partial history of the quotes and do not have the purchased coverage options. These are truncated to certain lengths to simulate making predictions with less history (higher uncertainty) or more history (lower uncertainty).	Free
Nielsen Data sets	http://research.chicagobooth.edu/nielsen/datasets/	Consumer Panel Data	The Consumer Panel Data include information about product purchases made by a panel of consumer households across all retail outlets in all US markets. The data include purchases from all Nielsen-tracked categories, including food, nonfood grocery items, health and beauty aids, and selects general merchandise. The data represent approximately 40,000 - 60,000 US households that continually provide information about the makeup of their households, the products they buy, as well as when and where they make purchases. The Consumer Panel Data comprise a representative panel of households that continually provide information about their purchases in a longitudinal study in which panelists stay on as long as they continue to	Free but only provided for study purposes so they must get confirm that the person who submit the request is registered in any University.

Data Set Owner	URL	Data Domain	Data Description	Available
			meet Nielsen's criteria. Nielsen Consumer panelists use in-home scanners to record all of their purchases (from any outlet) intended for personal, in-home use. Consumers provide information about their households and what products they buy, as well as when and where they make purchases. Years Available: 2004 - 2011.	
INFORMS Society for Marketing Science	https://www.informs.org/Community/ISMS/ISMS-Research-Datasets/Dataset-1	ISMS Durable Goods Dataset 1-Purchases history	There are 19,936 households who made 173,262 transactions involving durable goods purchases and related services from 1176 different stores of a major U.S. electronics chain. The transactions took place between December 1998 and November 2004.	$ 500
INFORMS Society for Marketing Science	https://www.informs.org/Community/ISMS/ISMS-Research-Datasets/Dataset-2	ISMS Durable Goods Dataset 2 - Response to promotion	This dataset records customer response to a Christmas promotion campaign offered by a major U.S. consumer electronics retailer. There are 176,961 customers in the database, 88,336 of whom were mailed the promotion; 88,625 of whom were not. Retail sales during the promotion period are available for both sets of customers. There are 152 variables for each observation, most of which represent each customer's purchase history before the promotion	$ 500
KDD CUP 2009	http://www.sigkdd.org/kdd-cup-2009-customer-relationship-prediction	Customer Relationship Prediction	This data was provided for the challenge of KDD cup 2009 The task is to estimate the churn, appetency and up-selling probability of customers; hence there are three target values to be predicted. The challenge is staged in phases to test the rapidity with which each team is able to produce results. A large number of variables (15,000) is made available for prediction. However, to engage participants having access to less computing power, a smaller version of the dataset with only 230 variables will be made available in the second part of the challenge	Free

AllState data set. [Data set clarification] could be considered as a good choice for using in research according to the following reasons

1. Available for free for academic purposes

2. It is a large data set as it contains 665,249 records for customers

3. This data set could be used in three dimensions of CRM: identification, attraction, and retention.

4. Data set format is CSV format, which facilitates the preprocessing and mining process.

In addition, researcher will use classification techniques over this data set, because classification fits to problem nature in which we use attributes of customer's data set to classify new customers to most suitable coverage options for them.

4. OBSTACLES AND CHALLENGES
The challenges in the CRM area so ramified. It can be classified according to the trends of CRM, from business view, or analytical view. This paper will present the challenges from two points of views: the business and analytical point of view. The following business obstacles facing the researcher locally:

1. The hardest challenge ever is to collecting a CRM data for the research. For any company the customer data is a highly sensitive data, even if the researches presented a documented proof that this data will be used for research purpose only. There are some banks didn't allow the analyst to access the customers transactions except in the case of churn high value customer.

2. It is a common challenge that the organizations also face with the traditional CRM or SCRM, that CRM is a business strategy not computer software. Most of organizations though that by installing software, all the customer issues should be solved. The Goals that the company need to achieve from the CRM or to be accomplished by CRM must be well known. The CRM objectives and philosophy must be spread around all the organization.

3. Companies expect from IT team to handle the CRM systems, while it need human resources from related departments such as sales, marketing and customer services to roll out the system after a sufficient and strict training

4. Technology is one of the greatest challenges for the SCRM, which social media to focus on. For example the selection of the suitable social media to use that will be valuable to the organization. In social media 78% of the consumers trust only the recommendations from social graph generated by like-minded individuals [12]. So it is a very important decision to choose the social media that have been used by most of the customers.

5. Trust a very hard challenge. Sometimes the researchers have to work with amount of data that may be less than the minimum required. As in the beginning of the customer life cycle with any organization, the customer was very conscious. So at the relationship between company and the customer start with a very little amount of data then increase over time.

The Challenges according to the Analytical CRM are as following:

1. There are a lot of unavoidably noisy data that we must deal with. Also the duplication of same data as every department deal with the customer may save the data by many different ways many times.

2. Non-trivial results almost always need a combination of different algorithms [5].

3. In CRM data came from different sources, so this need a strong integration before the analysis of these data. The different departments in the same organization capture and store the customer data in different sources that lead to a big waste in time to integrate the data. SCRM make this challenge harder as the data captured from different social networks, such as wikis, social bookmarking, online gaming, news aggregation …etc., which leads to a huge increase in the data sources and data amount that the company has to deal with.

Most of the current analytical models built based on the customer purchase pattern. These models do not have a deep understanding of customers which lead to a lot of wrong customer action predictions. Models based on the customer behavior should take more focus as a research point.

5. CONCLUSION

This paper has demonstrates the data mining techniques and tools that used in aCRM. In addition, it reviews categories, dimensions and components of any CRM. CRM new trends such as sCRM and iCRM have been declared as most of researches are going forwards towards the social media and the intelligent areas. The suitable data mining techniques to be used through each CRM dimension have been clearly identified within a detailed table that has been constructed from previous researches. CRM components have been clearly presented with its effects on the customer's behavior during the customer life cycle. This paper shows a list of free available customer's data set could be used in mining. Finally it lists the challenges of mining customer data from business and aCRM perspectives. For the future work the researchers are intended to conduct an experiment using different classification algorithms to achieve an enhancement in CRM's customer identification tasks. The classification algorithms more fit the research problem, as the available attributes in the data set make it a good base for CRM identification, attraction and retention. So the researcher decided to apply the research on the first task which is target customer analysis. This may be a good start to complete the whole CRM cycle on this data set in the future.

6. REFERENCES

[1] E. Peelen, Customer Relationship management. FT-Prentice Hall, 2005.

[2] E. Turban and Others, Electronic Commerce 2008 , A Managerial Prespective. Pearson Prentice Hal, 2008.

[3] Jiawei Han; Micheline Kamber, Data Mining: Concepts and Techniques. Elsevier, 2011.

[4] Adela Tudor , Adela Bara , Iuliana Botha, "Data Mining Algorithms and Techniques Research in CRM Systems," Recent Researches in computational Techniques non-linear Systems and Control, pp. 265-269, 2010.

[5] M. R. Farooqi and K. Raza, "a Comprehensive study of CRM Through Data Mining Techniques," in Preceedings of the National Conference ; NCCIST-2011, New Delhi, 2011.

[6] E.W.T. Ngai, Li Xiu, D.C.K. Chau, "Application of data mining techniques in customer relationship management:," Expert Systems with Applications, pp. 2592-2602, 2009.

[7] C. Rygielski, J.-C. Wang, and D. C. Yen, "Data Mining techniques for customer relationship management," Technology in Society, pp. 483-501, 2002.

[8] Carmen Petrisoaia, Nicolae Al. Pop., "The Role of Rational, Emotional and Spiritual Knowledge in Customer Relationship Management," in Proceedings of the 14th European Conference on Knowledge Management: ECKM 2013, 2013.

[9] Adrian Payne, Pennie Frow, Strategic Customer Management: Integrating Relationship Marketing and CRM. Cambridge University Press, 2013

[10] K. K. Wong, CRM in Action: Maximizing value through market segmentation, product differentiation & customer retention. iuniverse, 2011.

[11] Celina Olszak ; Tomasz Bartuś, "Multi-Agent Framework for Social Customer Relationship Management Systems," Issues in Informing Science and Information Technology, 2013.

[12] Fatimah Ayidh, Tanzila Saba, "Imapct of Social Networks on Customer Relation Management(CRM) on Prospectus of Business Environment," Journal of American Science, 2013.

[13] Carolyn Heller;Gautam Parasnis, "From social media to Social CRM : Reinventing the customer relationship," 2011.

[14] Z. Baracskai, V. Bevanda, and V. Dorfler, "Intelligent Customer Relationship managment (iCRM) by eFlow intelligent Portal," Intelligent information systems symposium, 2002.

[15] T. Saldanha and M. S. Krishnan, "Business Intelligence and CRM for Customer Involvement in Product and Service Development," in INFORMS Conference on Information Systems and Technology (CIST 2011), 2011.

[16] A. Faed, C. Wu, and E. Chang, "Intelligent CRM on the Cloud," in 13th International Conference on Network-Based Information Sydtems, 2010, pp. 216-223.

[17] A. Habul, A. Pilav-Velić, and E. Kremić, "Customer Relationship Management and Business Intelligence ," in Advances in Customer Relationship Management. www.intechopen.com, 2012.

[18] Gordon S.linoff, Michael J.A.Berry, Data Mining Techniques for Marketing, Sales & Customer Relationship Management . Jhon Wiley & Sons., 2011.