ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 2, Issue 7, July 2015

Improving Quality of Lead Data of Customer Relationship Management in M-Commerce

Kavya M Gowda[1], Dr. Vijaya Kumar B P[2]

Student, M. Tech (Software Engg,) Dept. of Information Science Engg, M S Ramaiah Institute of Technology

(Autonomous Institute Affiliated to VTU) Karnataka, India[1]

Professor& Head of Department, ISE, M S Ramaiah Institute of Technology, Karnataka, India[2]

Abstract: In the present growing trend of mobile computing, technologies, applications mainly focus on mobile e-commerce (m-commerce) and the mobile Web. As the mobile commerce market grows, Customer relationship management (CRM) is one of the major applications that incorporates the present marketing standard of relationship management (RM) and supports in acquisition, understanding requirements, and maintaining long-term relationships with customers. Expanding the CRM applications to mobile devices is increasingly becoming a major goal for organizations to optimize their workforce. To obtain a successful and desired outcome, quality of acquired data has a major impact on the organization's productivity. Hence the quality of data constitute the basis for major decisions of organization's operational and strategic levels which intern leads to positive growth of organizations. In order to tackle the above issue, we have proposed a novel quality model for customer lead data. Some of the activities are defined to analyse the performance of system with respect to time, data quality, reliability. This framework needs a feasible implementation module in the near future for improving quality of CRM lead data in m-commerce.

Keywords: CRM; Lead data; m-commerce; Data quality; e-commerce; mobile.

1.1 INTRODUCTION

In present global business environment, an advance in ecommerce technology has changed the way businesses is conducted across the globe and are mainly providing customers with greater convenience. With this new growing trend, a new set of challenges or issues are emerging mainly concentrating on mobile e-commerce (m-commerce).

Mobile e-commerce (m-commerce) is the concept that describes online business transactions that use wireless electronic devices such as hand-held computers, mobile phones or laptops, iPads, etc., These wireless devices interact with computer-mediated networks that have the ability to conduct online commercial business transactions. In recent years, many organizations have upgraded to mobile CRM to enhance their customer bases and loyalty. CRM applications extensively use lead information as base for crucial decisions in growth of business.

In Customer Relationship Management (CRM), a lead is an unqualified business opportunity, an initial interest and a potential relationship. Leads represent interested customers that sales representatives need to qualify to proceed or disqualify to stop further process. Lead data is easy to capture and the information acquired is used for further communication with potential customer. In addition, the term Data in real world is constantly changing, difficult to authenticate and cannot be used if it is becoming irrelevant and outdated. Huge data will increase risk of data maintenance. Manual auditing will lead to inefficient work which leads to failure. In this paper, we present a methodology to address the issues of bad lead data in CRM and we measure the results of the idea and analyse the results based on time, volume of bad. Data generated, increase in quality factor with respect to the huge volumes of lead data.The remainder of the paper is organized as follows: First section discusses some of the related work and the approaches used to improve the data quality, next literature on CRM and data quality is discussed in Section 3. Next, Section 4 proposes a model to improvise the quality of Lead data. Section 5 presents a case study to illustrate the application of the proposition. The paper ends with summary and conclusion in section 6.

1.2 RELATED WORK

There have been number of studies focusing on the quality of lead data in CRM. According to Madnick, Haug, Batini, poor quality data can have significantly negative impacts on the efficiency of an rganization, while high quality data are often crucial to a company's success. However, several industry expert surveys suggest approaches to improve quality of data. Some of them are as follows:

1.2.1 Eliminate duplicate records

An estimated 10 percent of names and addresses in an average CRM database are duplicate records. Identifying duplicates and merge/purging them are critical components to improving CRM data. But it is not as easy as it sounds. There is a number of software or technology based reduplication services which can help weed out duplicates and help merge multiple records into one for each customer which helps you gain valuable insight into customers to improve sales and marketing efforts. Some of the field matching algorithms [9] are used in elimination duplicates in databases are character based algorithms, q-gram-based algorithms, tokenbased algorithms [9] and so on.

1.2.2 Standardising

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 2, Issue 7, July 2015

Every organization should develop a quality standard that defines "bad"data in their database. The standard is really a collection of rules or tests that, when applied to the database, identify bad data and, in some cases, automatically fix it. After establishing an initial standard, team should continuously seek to improve and update it so that quality remains at a high level no matter how quickly the database evolves.

1.2.3 Data cleansing

Data cleansing, data cleaning or data scrubbing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database. Used mainly in databases, the term refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc[10]. parts of the data and then replacing, modifying, or deleting this dirty data or coarse data. Having defined a quality standard for their database, organizations must now implement it using data cleansing tools. With these tools. To maintain a consistently high level of data quality, (automated) searches and merges should be conducted on a set schedule perhaps daily or weekly.

1.2.4 Validation

Data validation stage is executed to check if the data is as required for further use. After the validation stage, data should be verified for further updation of records in the database to indeed conform to the quality standard.

1.2.5 Keep data up-to-date

Every year, approximately around 11 percent of customers and businesses relocate annually. As a result, data is constant and ever-changing, leading to outdated, inaccurate information a huge blow to the quality of your CRM database. "It's critical to have a consistent process for updating and managing customer data," says Holly Anderson, product marketing manager at K2, a provider of business process application software[7]. Generally it is very difficult to enter customer information every time they are working on a deal, but by providing an easy-to-use application, you can ensure that accurate data is collected every time and updated.

1.3 CRM

Customer Relationship Management (CRM)[11] is an application strategy used to help organizations to manage their customer relationship process. Customer data is one of the major concerns in order to provide successful results. The organization equipped with CRM software will have easy access to their customer data. With growing technologies and business strategies, the concept of mobile CRM has emerged as a buzz word. Innovations in mobile technology have opened opportunities to expand service features in business area as well [13].

Mobile CRM is a tool or application which has been emerged from wider field of CRM which utilizes mobile technology to collect and validate customer data over internet. It also provides statistical analysis of gathered data which enables organizations to perform more efficiently in CRM. Mobile CRM will support the remote employees to access, update and interact with customer data anywhere any time that helps to increase productivity of the company.

1.3.1 Leads in CRM

Leads facilitate company's opportunity to make business. Leads provide a streamlined link between marketing and sales, to accelerate the process between first interest and sales. Often Leads provide insufficient and incomplete details of customers, but the data would establish potential business for the organization. Hence, Lead management raises its significance in CRM area. Lead processing is a part of Lead management in CRM and it improves the interaction with customers and understanding of their needs. Leads are created automatically in the interaction center, campaigns, advertisements, or via HTML surveys on the Internet, road shows etc. The data collected during the Lead creation is termed as Lead data. Lead data plays an important role in the success or failure of CRM application intern refers to the success or failure of organization's business.

1.3.2. Data and its effects

Data are used in almost all the activities of organizations and hence can be termed as heart and soul of organizations and data constitute the basis for decision making on operational and strategic levels. Raw data describes the facts and figures that a company processes every day. In a business paradigm, every approach towards sale will be recorded. With close observation and analysis, the raw data becomes information. Analysis of daily sales information will reveal useful trends and patterns, such as the combination of products the customer buys, customer interests, customer visits, customers comfortable way of shopping online/offline, etc. Knowledge is a set of understanding based on the relationship between pieces of information.
Organizations analyse, collect and report on huge data. In 2007, Cindi Howson in her book on business intelligence, analysed that the average manager spend two hours per day hunting for data. About 50percent of the information located turns out to be useless. Organizations need to come up with efficient methods to turn their data into usable information to run their successful business. In the analysis and action, data is increasingly of major importance to businesses. Bad data or poor quality data can impact negatively on the growth of an organization in CRM, while high quality data are often crucial to a company's success. Today the knowledge and facts about customer data can chinch a deal. Consumers have more power in their transaction and organizations are competing for supremacy.

1.3.3 Data degradation

Data degradation is also termed as data corrosion. The data has been captured and stored correctly, but changes in the real world have rendered the data to be stored incorrectly. The term degradation or corrosion can be used when the data stored in a database, changes over time. Because of the ever changing world and nature of the

DOI 10.17148/IARJSET.2015.2708

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 2, Issue 7, July 2015

customer base, it is an obvious statement to say that the customer data does degrade over time.

Research on the data degradation says that not all data degrades over time. Let us consider an example of possible outcomes of data degradation in different areas like Business, Hospitals and Chemicals. The experiments on chemicals data did not degrade at all. In many cases, the results of an experiment made earlier were as important and useful as if the experiment and data survey had been done recently. So while some data degrades or corrodes, other data does not degrade or corrode at all. The experimental data on Hospitals showed the result in which the data is either accurate or not, and time plays a role with few interests of data such as age, marital status and so on, but whereas the blood group, DNA structures remains same. The data in the Hospital record is accurate over time for few fields, where as it degrades over time on other fields. So here is another database where data may or may not change depending on the domain of interest.

Here is another example of CRM customer database. In this customer database is recorded information about products, customers, distributions, sales, etc., found inside the company. The customer data keeps changing every day like age, interests and so on at this point, the data in the database keeps changing and hence it degrades.

From the above examples, we can observe that the degradation of data can occur or may not in a database. Some databases do corrode, some databases do not corrode. It cannot be guaranteed that the database corrodes at a constant rate. It is likely that the corrosion rate relates Poisson distribution.
The analysis on degradation of data helps to overcome these obstacles to make effective business decisions and success rates. The analysed result of data degradation can be used to estimate the eventual failure time the business.

1.3.4 Causes of data degradation

Some of the reasons for data degradation in CRM include:

- Contact/Location change: It is obvious to expect the customer location changes and contact details such as telephone numbers, e-mailIDs, address relocations and so on.
- Individual presence: Life status like alive and dead is one of the major reasons.
- Designation/Job changes: Designation or jobs change as a result of promotion, change in career, jobs etc.,
- Basic data changes: Basic data would include changes such as name changes, marital status change, family counts, etc.,

1.3.5 Effects of data degradation

The major challenges faced in the present world with data are that its quality degenerates over time. Experts say that most of the records in a customer database changes just in a span of one month because customers may change their contact numbers, address, may die, divorce, move etc. In recent survey most of the organizations lack in the initial planning to improve data quality. The problems in maintaining data will include from duplicate data, incorrect, incomplete data, etc[12].

The speed of data changes can explode the business over a short period of time. Hence, such changes to the data, requires companies to be capable enough to react instantly to the changing requirements and environmental conditions from customers. With negative impact, decisions grow more complex to compete in the present global business environment.
In the real world, majority of business organizations are aiming towards digitization of applications, new innovations and ideas to collect information and to use cost effective devices to bring into a new era where huge information can be gathered on all topics of interest to a business. Online transactions, social networks such as Facebook, WhatsApp etc., Mobile phones, and so on produce torrents of data in every hour as a by-product of their ordinary operations. Every individual is a data creator or can be termed as data source where the data is not in usable format. Effective analysis on the unstructured data helps in powerful decision making.

The most common effects caused by such bad data are the time and cost required to reconcile data and the loss of the system's or application's credibility. Sometimes there are chances that the employees avoid using the data because it is poor. The irony is that companies go to great lengths to manage their cash but not their equally valuable information.

1.4 APPROACHES TO IMPROVIZE DATA QUALITY

By implementing a strong customer data quality management program, organizations can begin to realize this vision of CRM. Many models have been proposed the improvisation of data quality having dimensions of quality, heterogeneous data as concern.

1.4.1 Proposed model

We are using a model to achieve a high quality data which is defined by the organizations and to achieve a high quality data that meets the quality success percentage which further decides whether to use the data in the business or to refine the data further as they have not met the quality significance level.
Here we propose a methodology to improve the data quality which intern helps the organizations growth.

The model in Figure 1 initially proceeds in tracking the changes triggered by the user or customer in the database via automated scripts. The next stage is to validate the triggered changes to check if the data is valid and to verify that the record is complete with all the mandatory data. The validation and verification is tested with the significance criteria set with each organization perspectives. Once the significance level is achieved, the synchronization of the databases is begun and all the delta changes tracked and validated are synchronized and updated with all the databases that the organization uses and this step also achieves in versioning for future data tracking uses.

DOI 10.17148/IARJSET.2015.2708

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 2, Issue 7, July 2015

Finally the overall testing is done to make sure all the data is successfully updated to the databases without any data loss due to network issues, server connection, and memory issues and so on. The procedure is iterated till the target success rate is achieved. The data is qualified only when the test success criteria is achieved, and hence is qualified as good data that can be used in the business decisions.

Stage I: Periodic automated scripts to track the changes: In this stage periodic automated scripts implemented are executed to schedule the jobs based on the organizations requirement, like scheduling hourly, daily, weekly, monthly or on file drop. The automated scripts track all the changes triggered by the user or customer and also record the time stamps and the sources like who triggered the change? What change was triggered? When wad the change triggered? And so on.

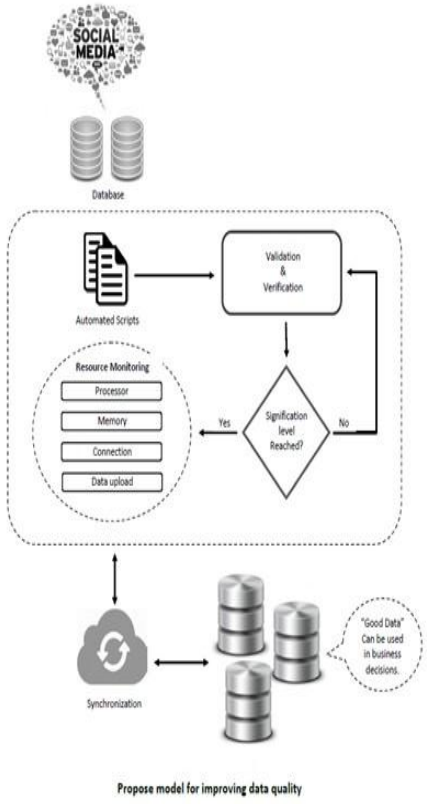

Fig.1. Proposed model for improving data quality.

Stage II: Validation and Authentication of data: At this stage of validation, the organization identifies and corrects the errors in the changed data. Misspellings, transpositions and other anomalies can be mended, as well. Data validation will mainly look in particular for the data quality issues like redundancy check, incorrect data populated, data mappings, search and sort functionality, missing data, graphs representation, unnecessary dependencies, etc.

Qualification Level code	Description
3	Bad data
6	Inconsistent data
9	Good data

Fig. 2. Organization specific significance level codes.

The validation and verification criterion is tested based on the significance level codes defined by the organization as shown in Figure 2. If the data fails to qualify the significance level, the process is re-triggered and fine-tuned until the significance is achieved. On successful results of this stage, the process is preceded to further stage of synchronization and versifying.

Example: Lets us consider the company X fixes the significance level as 9 in order to qualify the data, and then all the records which fail to achieve the significance level will be pushed for previous processing until the data comes out with the qualifying significance level.

Stage III: Resource monitoring: This stage of analysing resource availability checks if the required memory space is available as per the internal calculations which give out output of the expected results, connection are configured, no network issues persists, etc., once the resource monitoring activity results in positive, the synchronization activity is triggered for further uploads of data to the respected databases.

If the synchronization fails due to any dynamic changes in network or connection problems, the percentage of data that is uploaded to databases is tracked and the remaining data is pushed again to the loop of resource monitoring phase.

Stage IV: Synchronization of the databases and versioning: This stage initializes synchronizing jobs when the resource monitoring stage is achieved. Synchronization testing covers the following main areas and scenarios:

- New and updated data saves properly.
- Synchronization will not disturb the regular functionality.
- Updating data is smooth and does not require you to move to another form.

Stage V: Versioning: Once the data is uploaded to the databases successfully, it is versioned for future use to track the history of changes.

1.5 RESULTS AND ANALYSIS

As of 2015, Exabytes of data are created each day, and that number is doubling every 20 months or so. More data cross the internet every second than were stored in the entire internet just 20 years ago. This gives organizations an opportunity to work with many Exabytes of data in a single data set and not just from the internet. For instance, it is estimated that Walmart collects more than 2.5 petabytes of data every hour from its customer transactions. A petabyte is one quadrillion bytes, or the equivalent of about 20 million filing cabinets' worth of text. An Exabyte is 1,000 times that amount, or one billion gigabytes.

Let us consider an example, where organization X uses the CRM lead data in order to take business decisions for the growth of the organization. We assume the organizations handles lakhs of records containing customer data which would degrade over time. Here our objective is to track the changing data and validate it to qualify the data to the business decisions.

DOI 10.17148/IARJSET.2015.2708

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 2, Issue 7, July 2015

In the Figure 4, graphical representation shows the analysis of data qualification based on the significance level code. Here, the table describes the significance of code. The signification code definition varies from one organization to other with respect to the organizational standards. Here in the graph, the X-axis depicts the time stamp of the automated script executions and the Y-axis depicts the number of records (in lakhs).

In the graph, the first bar at time stamp 1 has 1 lakh records of good data, 3 lakh records of inconsistent data and 2 lakh records of good data(which have passed the significance level "9"defined by the organization). Hence, at time stamp of 1 hour, we have 2 lakh good records which can be used in further business processes and decision whereas, the remaining records are pushed back in loop for processing until they meet the organization's significance level.

Organization specific significance level codes

Qualification Level code	Description
3	Bad data
6	Inconsistent data
9	Good data

Fig. 3. Organization specific significance level codes.

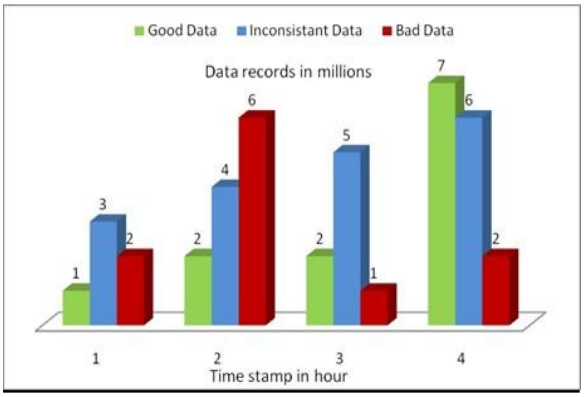

Fig. 4. Graphical analysis of data quality versus time stamp.

In our model, we assume that organization maintains a significance level code as "9"in order to test the data validity. Hence, the periodic automated scripts are implemented to track the source of changes and changes triggered to the databases, and the changed records are subjected to validation in order to remove the inconsistent data and the validated data is tested to check for the significance level defined as shown in Figure 3. The records which achieve the significance level are qualified and are subjected to further process of synchronization, where as the records that do not meet the criteria are subjected to the iterative process of validation till the records meet the significance criteria.

Once the synchronization is achieved successfully, the data is termed as "Good data" and can be used in further CRM decisions. If the synchronization is not achieved successfully, only the success percentage of data is uploaded to the databases and remaining data are again subjected to the synchronization. The process iterates till 100 percent data upload is achieved and hence the reliability of the data is increased comparatively.

1.5.1 Measurements and analysis

Some of the performance analysis has been observed taking general example and the implied results are shown with graphical representation in this section.

Bad data generation

As the number of data records increases at every time stamp, the bad data generation also increases. and the time required to process the bad data also increases.

In the Figure 5, the X-axis depicts the number of records in lakhs and Y-axis depicts the time stamp at which the automated scripts are subjected to execution (based on organization's standards), the percentage of bad data generated increases with the increase in the actual data subjected to processing.

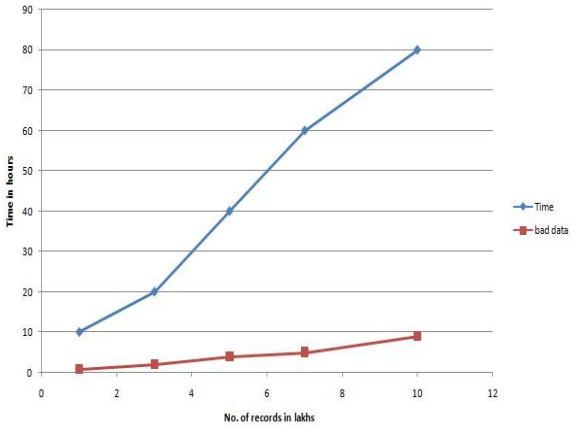

Fig.5.Bad data generation.

Version delay

Let us assume we have 10 lakh total records which are subjected to synchronization and the outcome would have 2 lakh records are perfectly synchronized whereas remaining 8 lakh records have to be pushed back for synchronization. Hence, the we have time delay of versioning till all the remaining records achieve 100 percent success. The graph below in figure 6 shows the implied results of version delay expected.

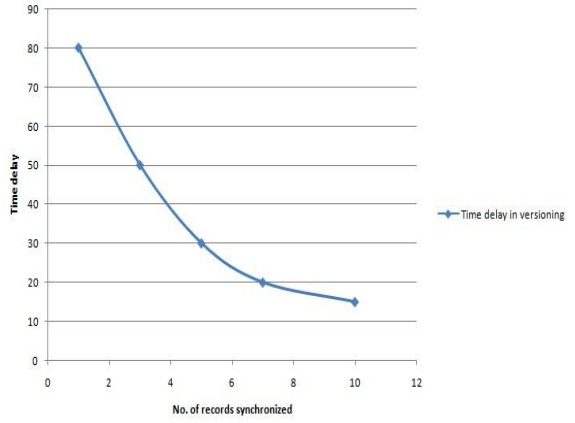

Fig.6.Time delay in versioning.

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 2, Issue 7, July 2015

In the Figure6, the X-axis depicts the total number of records synchronized at a particular time stamp and Y-axis depicts the time delay expected in achieving 100% synchronization. As the time delay increases the time to version the synchronized data also increases as shown in the graph.

1.6 CONCLUSION AND FUTURE WORK

This section gives the summary of our key observations and conclusions involves results based on the analysis of social media data of existing systems and are made to work towards practical implementation on how the results can be verified in our future work.

By following the proposed approach for obtaining the significant level of data quality maintenance efforts, we find that the optimum level of data is achieved by meeting the significant percentage set as per the organization criteria. The significance level can be achieved by following a approach of periodic automated scripts running for tracking the changed data with versioning capability, Validation of the changed data Synchronizing the databases with the changed data (delta change) and Tests to achieve the significance level.

As the model shows, the optimal level of data quality maintenance is not to achieve perfect data at low cost, but to avoid the poor data and to guarantee the quality of data. This data maintenance effort is dependent on the characteristics of the particular company. Different industries have different characteristics, different significance, different target criteria to be met i.e. the relation between improvised data quality and assuring data quality.

To sum up, this paper has produced a better understanding of how to define the optimal data quality and although these contributions are to be further elaborated on in future research, in their present form they provide a better understanding of the topic which hopefully aids companies in their data quality work.

ACKNOWLEDGEMENT

I am indebted to **Mr. Kushal Keshav** for his valuable suggestions and excellent support in completing this work at right time and also thanks to management of MSRIT, Bengaluru for providing the academic and research atmosphere at the institute.

REFERENCES

[1] Tenopir, C.,"Database quality revisited", Library Journal, 1 October 1990, pp. 64-67.
[2] Wang, R., Kon, H. and S. Madnick, "Data quality requirements analysis and modeling", 9th International Conference on Data Engineering, 1993, pp. 670-677.
[3] Wang, R., Strong, D. and L. Guarascio, "Beyond accuracy: What data quality means to data consumers", Journal of Management Information Systems, Spring 1996, Vol. 12, No. 4, pp. 5-34.
[4] Redman, T.C., "Improve data quality for competitive advantage", Sloan Management Review, Winter, 1995, Vol. 36, No. 2, pp. 99-107.
[5] Costin, H., Total Quality Management, Dryden, United States, 1994.
[6] K. Tripathi and S. K. Nair, Mobile advertising in capacitated wireless networks, IEEE Transactions on Knowledge and Data Engineering, vol. 18, ch. 9, pp. 1284-1296, 2006.
[7] http://www.data-driven marketing.net/articles/organization/ three-ways-improve-your-crm-investment
[8] S. J. Barnes, The mobile commerce value chain: analysis and future developments, International Journal of Information Management, vol. 22, no, 2, pp. 91-108, 2002.
[9] Mingzhen Wei1, Andrew H. Sung 2, Martha E. Cather "Improving database quality through eliminating duplicate records", Data Science Journal, Volume 5, 19 October 2006 127.
[10] http//www.informit.com/articles/article.aspx?p=399325&seqNum=4
[11] Sarmad Alshawi, Farouk Missi, Zahir Irani "Organisational, technical and data quality factors in CRM, adoption SMEs perspective" Brunel Business School, Brunel University, UK.
[12] Ranjit Singh, Dr. Kawaljeet Singh "A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing" IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 2, May 2010 .
[13] Choi Sang Long, Raha Khalafinezhad, Wan Khairuzzaman Wan Ismail, Siti Zaleha Abd Rasid "Impact of CRM Factors on Customer Satisfaction and Loyalty" Asian Social Science; Vol. 9, No. 10;2013