iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VI **Month of publication:** June 2020

DOI: http://doi.org/10.22214/ijraset.2020.6230

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

Andromeda

Sahil Parashar[1], Divyank Pandey[2], Abhishek Singh[3], Anant Singh[4], Abhijeet Singh[5]
[1, 2, 3, 4, 5]*Apex Institute of Technology, Chandigarh University*

Abstract: Currently, we live in an analytical era_ everyone wants to increase their business revenue and also wants to improve their relationships with their customers. Because they know Customer service plays a major role in their business. It's the leading indicator for measuring customer loyalty. Identify your unhappy customers, reduce churn, and increase revenue. It's also a key point of differentiation which helps to attract fresh customers in a competitive business environment.
A golden record is a single, well-defined version of all the data entities in an organizational ecosystem. In this contest. a golden record is sometimes called the "single version of the truth," where "truth" is understood to mean the reference to which data users can turn when they want to ensure that they have the correct version of a piece of information The golden record encompasses all the data in every system of record within a particular organization.
MIMI (Master Data Management) it a comprehensive method of enabling an enterprise to link all of its critical data to one file, called a master file, that provides a common point of reference. When properly done, master data management streamlines data sharing among personnel and departments. In addition, master data management can facilitate computing in multiple system architectures, platforms And applications.
At its core MDNI can he viewed as a "discipline for specialized quality improvement" defined by the policies and procedures put in place by a data governance organization.
The ultimate goal being to provide the end user community with a "trusted single version of the truth" from which to base decisions.
Index Terms: Data Quality, Customer Relationship Management, Master Data Management, Golden Data, CRAM model.

I. INTRODUCTION

The existence of data alone does not guarantee that all functions and decisions of data management can be made. The primary purpose of data quality is about shucking data -missing or inaccurate or inaccurate data from a particular perspective. Most of the time, when data quality is achieved when an enterprise uses complete, understandable and harmonious data, inputting data quality is the first step in making data quality efficient and capable of understanding effectively and efficiently the data, the data must satisfy a set of quality criteria.
Quality policy satisfaction data is said to be of high quality. Ample efforts have been made to differentiate data quality and to determine its size. Dimensions of data quality usually include accuracy, reliability, and relevance. Consistency, accuracy, timeliness, precision, understanding ability. Consistency and utility. For our research work we have identified a quality indicator by taking 9 key items as outlined below. According to English 131 we describe the following aspects of the following information quality and measures.

1) *Definition Conformance:* The chosen Object is of most important and its definition should have complete details and meaning of the real world object.
2) *Completeness (of values):* Is the characteristic of having all required values for the data fields.
3) *Validity (Business rule conformance):* Is a measure of degree of conformance of data values to domain and business rules. This includes Domain values. Ranges. reasonability tests. Primary key uniqueness, Referential Integrity.
4) *Accuracy and the Source):* Is a measure of the degree to which data agrees with data contained in an original source.
5) *Precision:* The domain value which specifics business should have correct prevision.% as per specifications.
6) *Non-duplication (of occurrences):* Ls the degree to which there is a one-to-one correlation between records and the real world object or events being represented.
7) *Derivation Integrity:* Is the correctness with which two or more pieces of data are combined to create new data
8) *Accessibility:* Is the characteristic of being able to access data on demand.
9) *Timeliness:* Is the relative availability of data to support a given process .

II. METHODOLOGY

Data quality assurance is a complex issue that requires a systematic approach. English [3] proposes a comprehensive Data Management System, consisting of 5 steps to measure and improve information quality, and an umbrella process for bringing about cultural and environmental changes to reinforce information quality development as a management tool and practice:

1) Step I: Assess Data Definition & Information Architecture Quality
2) Step 2: Assess Information Quality
3) Step 3: Measure Non quality Information Costs
4) Step 4: Re engineer and Cleanse Data •
5) Step 5: Improve Information Process Quality •
6) Step 6: Establish the Information Quality Environment

Fig-I describes the steps of data quality to be accomplished to have proper and free from data quality factors which are described in the coming sections [2].

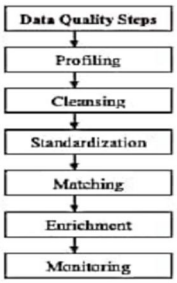


Fig.. I. Data Quality Steps

A. *Data Quality Steps*

The main goal of a Data Quality solution is to aggregate data from one or more data sources. However. the data aggregation process often results in a wider range of data quality issues that need to be addressed. For example, incomplete or missing customer prattle information may be disclosed, such as empty phone numbers or addresses. Or some data may be inaccurate, such as a customer record showing you live in the city of Wisconsin, in the Green Bay region. Fig1 describes 6 functions of data quality in a database [2]

1) *Profiling:* As the first line of defense for your data integration solution, data encryption helps you check if your existing data sources meet the quality standards of your solution. Properly printing your data saves time in practice because it identities issues that need immediate question from the start and prevents unnecessary processing of unacceptable data sources. The data profile becomes very sensitive when dealing with raw data sources that do not have the integrity of comments or quality control. There are many functions to filler data: column statistics, number distribution and pattern distribution. These functions analyze individual and multiple columns to determine relationships between columns and tables. The purpose of these data profiling operations is to enhance the clear image content of your data[2]
2) *Column Statistics:* This task identifies problems in your data, such as invalid dates. It reports average, minimum, maximum statistics for numeric columns.
3) *Value Distribution:* Identifies all values in each selected column and reports normal and outlier values in a column. Pattern Distribution Identifies invalid strings or irregular expressions in your data.

B. *Cleansing*

After the data set has successfully met the comment standards, it still requires cleaning and data provisioning to ensure that all business rules are properly met. Effective data cleansing requires the use of flexible and efficient methods that can handle complex quality issues that are hidden deep within large data sets.

C. *Matching*

Data matching consolidates data records into identifiable groups and links/merges related recon within or across data sets. This process locates matches in any combination of over 35 different components – from common ones like address, city, state, ZIP•, name, and phone – to other not-so-common elements like email address, company, gender and social security number.

D. Enrichment

Data enrichment enhances the value of customer data by attaching additional pieces of data from other sources, including decoding, demographic data, full-name parsing and tenderizing, phone number verification, and email validation. The process provides a better understanding of your customer data because it reveals buyer behavior and loyalty potential [2]

1) *Address Verification.*: Verify India addresses to the highest level of accuracy. To join all the people having same address as a family group.
2) *Phone Wide:* firm, Fill in missing area codes, and update and correct area code/prefix. Also Append lat/long, time zone, city. state, ZIP, and county. And Validate the phone numbers by using Twillio api.
3) *Email Validation*: Validate, correct and clean up email ad-dresses using three levels of verification: Syntax; Local Database; and MX1ookup. Check for general format syntax errors, domain name changes, improper email format for common domains lie. Hotmail, Gmail, Yahoo) and validate the domain against a database of good and bad addresses, as well as verify the domain name exists through the Mailed-x-change (MX) Lookup, and parse email addresses into various components.
4) *Name Parsing and Gendering. Parse:* Full names into com-opponents and determine the gender of the first name.
5) *Residential Business Delivery*: Indicator Identify the delivery type as residential or business.

E. The relationship with Customer

The term "customer lifecycle" refers to the stages in the relationship between a customer and a business. It is important to understand customer lifecycle because it relates directly to customer revenue and customer profitability. Marketers say there are three ways to increase a customer's value: (I) increase their use (or purchases) of products they already have; (2) sell them more or higher-margin products; and (3) keep the customers for a longer period of time [5]

However, the customer relationship changes over time, evolving as the business and the customer learn mom about each other So why is the customer lifecycle important? Simply put, it is a framework for understanding customer behavior. In general, there are four key stages in the customer lifecycle:

1) Prospects—people who are not yet customers but am in the target market
2) Responders—prospects who show an interest in a product or service
3) Active Customers—people who are currently using the product or service
4) Former Customers—may be "bad" customers who did not pay their bills or who incurred high costs; those who are not appropriate customers because they are no longer part of the target market; or those who may have shifted their purchases to competing products.

The customer lifecycle provides a good framework for applying data mining to CRM. On the "input" side of data mining, the customer lifecycle tells what information is available. On the "output" side, the customer lifecycle tells what is likely to be interesting [5]

Looking first at the input side, there is relatively little information about prospects except what is learned through data purchased from outside sources_ There are two exceptions: one, there are more prospecting data warehouses in various industries that track acquisition campaigns directed at prospects; two, click-stream information is available about prospects' behavior on some websites. Data mining can predict the profitability of prospects as they become active customers, how long they will be active customers, and how likely they are to leave [5]

In addition, data mining can be used over a period of time to predict changes in details. It will not be an accurate predictor of when most lifecycle events occur. Rather, it will help the organization identify patterns in their customer data that are predictive. For example, u firm could use data mining to predict the behavior surrounding a particular lifecycle event (e.g., retirement) and find other people in similar life stages and determine which customers are following similar behavior patterns [5]

The outcome of this process is marketing data intelligence, which is defined as "Combining data driven marketing and technology to increase the knowledge and understanding of customers, products and transactional data to improve strategic decision making and tactical marketing activity, delivering the CRM challenge" [6] There are two critical components of marketing data intelligence: customer data transformation and customer knowledge discovery. Raw data extracted and trans-formed from a wide array of internal and external databases, marts or warehouses and the collecting of that total data into a centralized place where it can be accessed and explored is data transformation. The process is continued through customer knowledge discovery, where the information is mined, and usable patterns and inferences can be drawn from the data. The process must be measured and tracked to ensure that the results fed to campaign management software produce information that the models created by data mining software find useful and accurate [6].

Data mining plays a critical role in the oven-all CRM process, which includes interaction with the data mart or warehouse in one direction, and interaction with campaign management software in the other direction_ In the past, the link between data mining software and campaign management software was mostly manual. It required that physical copies of the scoring from data models be created and transferred to the database. This separation of data mining and campaign management software introduced considerable inefficiency and was prone
to human error. Today the trend is to integrate the two components in order to gain a competitive advantage [7]. Firms can gain a competitive advantage by ensuring that their data mining software and campaign management soft-ware share the same definition of the customer segment in order to avoid modeling the entire database. For instance, if the ideal segment is high-income males between the ages of 25 and 35 living in the northeast, the analysis should be restricted to just those characteristics. In addition, the selected scores from the predictive model should flow directly into the campaign segment in order to form targets with the highest profit potential[7].

F. Construction of Model

CHAID 141 is simple and fast to build, and neural networks have many parameters to set up and require more sophisticated manipulation to make sure they are the best fit It is difficult to apply background information using neural networks, and it is easy to detect errors and adequacy in u CHAID tree.

More data set should be given to the CHAID model to ensure that there is a critical mass in leaf-shaped areas following multiple branches. The details of both methods need to be considered in advance. Neural networks require data to be converted to binary format Before using CHAID, any continuous independent variable must be banded.

1) Using CHAID Model to predict the &times' Idea' After preparation of data we can use our model to predict the new business ideas in the Area just by entering area pin code and the list of business ideas will be given to user having good chances to survive and get successful.

III. CONCLUSION

Choosing between different options is not as critical as choosing to use Data quality assurance technology in a CRM system. Data quality assurance represents connectivity from data stored over many years through multiple interactions with customers in a variety of contexts, and information needed to succeed in relationship marketing meetings. To unlock the potential of this information, data mining enables processing that can be extremely complex and time-consuming, and ex-tends to previously unknown factors used to improve customer retention, response rates, attractiveness, and sales volumes. With the full use of the CRM system. which should include data quality assurance, organizations arc promoting improved loyalty, increasing their customer base, and attracting the right customers.

As customers and businesses meet regularly, businesses will need to leverage on CRM and related technologies to capture and analyze large amounts of customer data Businesses that use customer data and personal information resources more effectively will be mom likely to succeed. However, businesses should also remember that they must use technology effectively to achieve a balance between privacy rights and economic benefits.

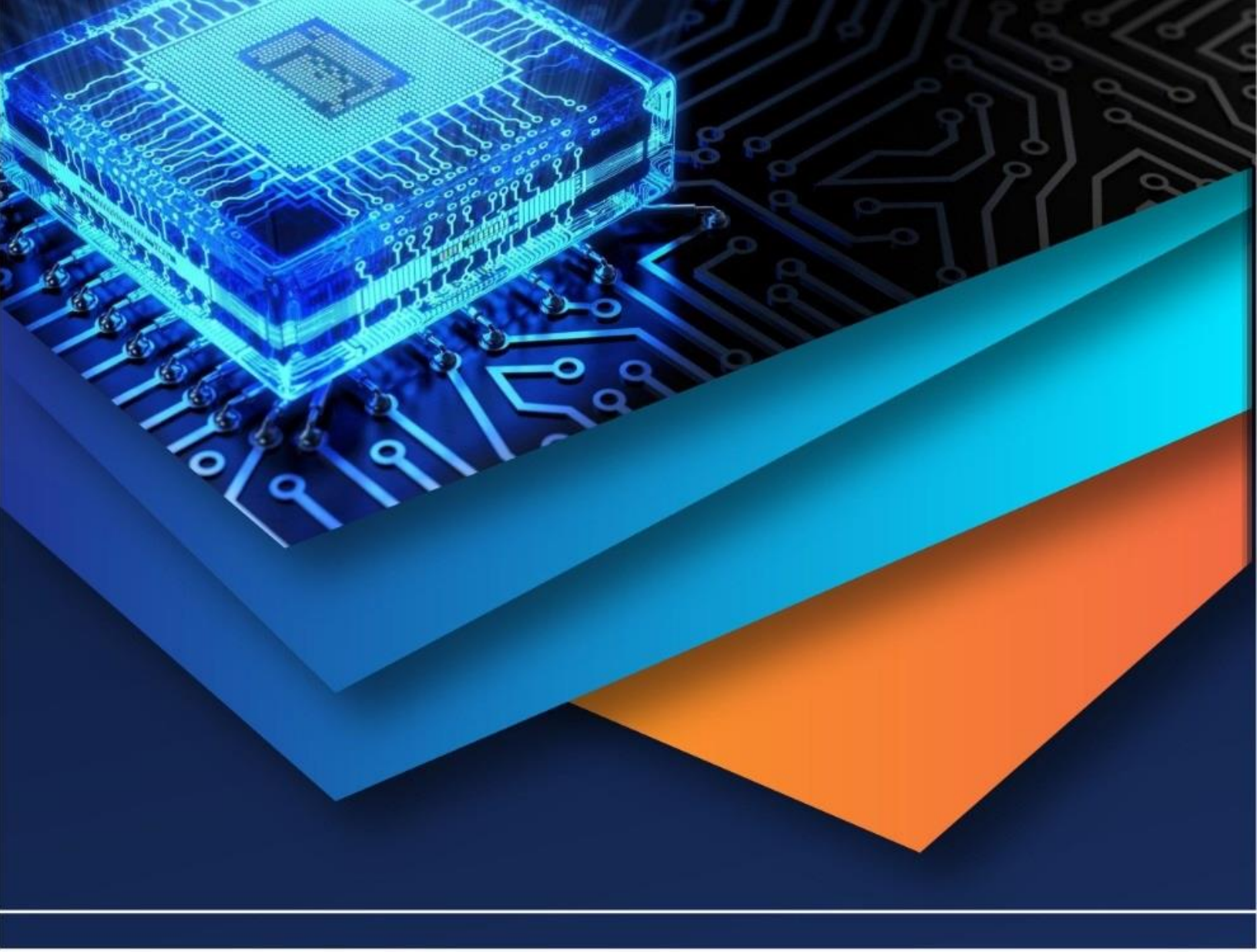
IV. FUTURE WORK

We can implement a platform independent app to handle all the company data and coven it to bogus free data.

We aimed to carry out work leading to the development of a GUI based platform independent Desktop App where the local shopkeeper. local marts, can verify their users details, manage their Customer database and can do analysis on that database.

REFERENCES

[1] Symbiotic Cycles of Data Profiling Integration. and Quality (poster) TDWI, 2006"
[2] Six Steps to Managing Data Quality with SQL Server Integration Services: www.MelissaDatacom. •
[3] English, L P. (1999). Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits, John Wiley and Sons, Inc. Data Quality Issues, Page10
[4] Applied Metrix Hcbsite. http://www.applicdnietrix.co m/CaseStud ic stbank.htm I
[5] M. Freeman The 2 customer lifecycles Intelligent Enterprise, 2 (16) (1999), p. 9
[6] Chablo E, Marketing Director, smartFOCUS Limited. The importance of marketing data intelligence in deliver-ing successful CRM. 1999.
[7] Thearling K, Exchange Applications. Inc. Increasing customer value by integrating data mining and campaign management software. Exchange Applications White Pa-per, 1998.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY