RELATIVE PARAMETER QUANTIFICATION IN DATA MINING – A CASE STUDY ON TELECOM CELLULAR MOBILE SERVICE PROVIDERS IN TERMS OF QOS IN INDIA

Mahesh Kandakatla[1], Prashanth Bolukonda[2] and Lokanatha C. Reddy[3]

[1]Research Scholar, Dept. of CSE, OPJS University, Rajasthan, India
[2]Assistant Professor, Dept. of CSE, Vaagdevi College of Engineering, Telangana, India
[3]Professor, Dept. of CS, School of Science & Technology,
Dravidian University, Kuppam, A.P., India

ABSTRACT

Interpreting available data is a focal issue in data mining. Gathering of primary data is a difficult and expensive affair for assessing the trends for any business decision especially when multiple players are present. There is no uniform formula-type work procedure to deduce information from a vast data set especially if the data formats in the secondary sources are not uniform and need enormous cleansing to mend the data for statistical analysis. In this paper, an incremental approach to cleanse data using a simple yet extended procedure is presented and it is shown how to deduce conclusions to facilitate business decisions. Freely available Indian Telecom Industry's data over a year is used to illustrate this process. It is shown how to conclude the superiority of one telecom service provider over the others comparing different parameters like network availability, customer service quality etc. using a relative parameter quantification technique. It is found that this method is computationally less costly than the other known methods.

KEYWORDS

Quantification, Data mining, QoS, Data cleansing.

1. INTRODUCTION

Data represents information in the form of facts or entity instances. Data is part of miniworld(M) or Universe of Discourse(UoD)[1]. Maintaining of quality in large datasets is an issue because of many reasons. Missing values appear frequently in real world and business related databases like companies, governments and academia for many different reasons[2][3][4][5]. Poor data quality is an issue and dealing with it is a challenge.

Joe F. Hair JrKennesaw stated that traditional researchers criticize data mining and predictive analytics as unscientific because the techniques are data driven instead of theory driven. But looking beyond these issues most researchers agree that data mining and predictive analytics have many advantages. Among the most important advantages is the tools identify relationships that would remain hidden, and decision-making is therefore more informed and better. The reason for

improved decision-making is more complete understanding of the data and its underlying patterns, and more accurate predictions. As a result decisions are more consistent and thus more accurate. When researchers and managers increase the accuracy of their decision-making the outcome is increased customer satisfaction and retention for the organization, and therefore lower costs and higher profits [6].

Business competition demands quick and rational analysis done on certain specific views of the data. The gaps between application software and modern storage gadgets with strong retrieval systems, for making effective interpretation demands force us to invent short methods even if a little of brute force is used – the effectiveness lies in minimizing the multiple scans of the same data. Relative quantification of decision parameters speeds up decision-making. Quantifiable business benefits have been proven through the integration of data mining with current information systems, and new products are on the horizon that will bring this integration to an even wider audience of users [7].

As such there is a need to devise simple procedures to form part of a larger decision making tool. A particular analysis on publicly available database to determine how to select the best player, for example, in telecom sector is expected to give boost to build small and simple tools in this direction. Motivation for making the present case study is from the data on Indian telecom operators'[8] facts pertaining to multiple quarters with different performance indicators. Single quarter analysis is avoided because there is a statistical likelihood of influence of natural or other calamities that may hamper availability of correct data components at times. Hence, a periodic set of data spread over a year with four quarters is considered and a variable decision parameter value to interpret the potential outcomes of study is analyzed.

2. MATHEMATICAL BACKGROUND

When real world data sets are studied, it is observed that a few components' values do miss in the tuples sometimes. The missing can be at any locations in an n-tuple. Similarly, a few n-tuples might miss from getting recorded properly. There are many statistical procedures to predict the missing values. Extensive research took place on this type of issues ([9], [10], [11], [12] etc). However, employing such techniques [13] in an interactive mode is prohibitively time consuming and will over run the time available for decision-making. Hence, it is tried to fill the missing values in a scenario based quick decision mode using elementary concepts paving way for Relative Parameter Quantification [14].

Berger [1985] stated that Decision problems present a difficulty in determining the best decision because an action that is best under one state of nature is not necessarily the best under the other states of nature. Although, various schemes have been proposed - decision principles that lead to the selection of one or more actions as "best" according to the principle used – none is universally accepted [15].

Lindgren (1971), French and Insua (2000) stated that ordering of available actions linearly, assigning "values" to each action according to its desirability is a frequentist principle. The minimax principle places a value on each action according to the worst that can happen with that action. For each action a, the maximum loss over the various possible states of nature:

M(a) =maxl(θ ,a)

is determined and provides an ordering among the possible actions [16][17]. Taking the action a for which the maximum loss M(a) is a minimum lends itself to the name minimax.

MiniMax Principle

In decision theory it is a principle for decision-making by which, when presented with two various and conflicting strategies, one should, by the use of logic, determine and use the strategy that will minimize the maximum losses that could occur. This financial and business strategy strives to attain results that will cause the least amount of regret, should the strategy fail [18] [19].

Essentially, try to choose the least of losses if we have to take a decision, which might be bad.

MaxiMin Principle

In decision theory it is a pessimistic (conservative) decision making rule under conditions of uncertainty. It states that the decision maker should select the course of action whose worst (maximum) loss is better than the least (minimum) loss of all other courses of action possible in given circumstances. Also called maximin regret or minimax criterion [18] [19].

Essentially, try to choose the biggest of gains if we have to take a decision, which might be good.

3. PROPOSED WORK

The Telecom Database [8] from TRAI(Telecom Regulatory Authority of India) made available the Telecom Services Performance Indicators PDF files in quarterly mode. These PDF files are mainly reliant on the data from the telecom cellular operators (Service Providers) from variety of sources like subscription of data, revenue and usage financial data of telecom service sector, QoS of wireless, Wireline and Dial-up/Broadband services and performance of cable TV, DTH and broadcasting services.

Data for this study is collected from Annexure 4.1 i.e. Performance of QoS parameters for cellular mobile service. Four quarterly PDF files of annexure 4.1 are transformed in to parallel excel files. Each of these files has data related to 17 parameters, which are occurring as 17-tuples with qualifiers.

The List of parameters considered for QoS is as under:

Table: 1 Parameters under Consideration

Parameter No.	Parameter Description	Performance Indicators (Should be)
P-1	BTSs Accumulated downtime (not available for service) (%age)	$\leq 2\%$
P-2	Worst affected BTSs due to downtime (%age)	$\leq 2\%$
P-3	Call Set-up Success Rate (within licensee's own network)	$\geq 95\%$
P-4	SDCCH/ Paging Chl. Congestion(%age)	$\leq 1\%$
P-5	TCH Congestion (%age)	$\leq 2\%$
P-6	Call Drop Rate (%age)	$\leq 2\%$
P-7	Worst affected cells having more than 3% TCH drop (call drop) rate (%age)	$\leq 3\%$
P-8	Connection with good voice quality	$\geq 95\%$
P-9	Point of Interconnection (POI) Congestion (No. of POIs not meeting the benchmark) (Averaged over a period of quarter)	$\leq 0.5\%$
P-10	Metering and billing credibility - post paid	$\leq 0.1\%$
P-11	Metering and billing credibility - pre paid	$\leq 0.1\%$
P-12	Resolution of billing/charging/validity complaints	100%
P-13	Period of applying credit/ waiver/ adjustment to customer's account from the date of resolution of complaints	100%
P-14	Accessibility of call center/ customer care	$\geq 95\%$
P-15	Percentage of calls answered by the operators (voice to voice) within 60 seconds	$\geq 90\%$
P-16	%age requests for Termination / Closure of service complied 1.00	100%
P-17	Time taken for refund of deposits after closures	100%

A single Excel file is generated in a single pass on these files with a segregated data organization so that each parameter has its own contiguous columns of the four quarters and rows of the different operators spread across different States/Territories of India. Incomplete tuples are also recorded with NULL values in this pass. Filling up the empty cells is done using Relative Parmeter Quantification.

Relative Quantification Parametrization

Relative quantification refers to the estimation of the changes in steady state observations. This type of change can be assumed as a fixed percentage of the value of the neighbour without loss of generality. We wish to calibrate this by a parameter following *Relative Parameter Quantification*. This type of estimates have been successfully carried out in DNA & RNA (Deoxy Ribo Nucleic Acid and Ribo Nucleic Acid) and other life-sciences related studies.

A comprehensive Scheme is furnished as Algorithm-A using this concept.

Algorithm – A: Single Scan File Generation

Input: Raw data is in Q_1,Q_2,Q_3,Q_4 excel files
Output: Regrouped data in file R(excel file)
Procedure:-
for each operator o do
   begin
      for each state s do
   begin

```
    for each data_group i do
              begin
                    q₁ᵢ=Q₁(relevant cell)
q₂ᵢ=Q₂(relevant cell)
                    q₃ᵢ=Q₃(relevant cell )
                      q₄ᵢ=Q₄(relevant cell)
    end
    begin_new_horizontal_record in R
              record(s,o)
    for each data_group i do
               begin
               record(q₁ᵢ,q₂ᵢ,q₃ᵢ,q₄ᵢ)
               end
               endfor_data_group_loop
    end_new_horizontal_record
        end
    endfor_state_loop
    end
endfor_operator_loop
```

A second pass on the output Excel file is done to cleanse the data using the MiniMax and MaxiMin procedures depending on the nature of each of the 17 parameters.

A filtering level at 75% of the MiniMax and 75% of MaxiMin cleaning method is taken in this case study. This completes the cleansing process. However, the filtering level 'α' can be made a variable and computation can be carried out quickly with different levels interactively if needed. A comprehensive procedure is given in Algorithm-B that uses available Excel-cell values to guess replacing value for NULL cell (for the MiniMax case). Algorithm-C is similar and not stated here.

Algorithm – B: Cleansing of Data based on MiniMax

Assumption: α is a variable global relative quantifier.

```
α = 0.75  /* i.e., 75% --- can be an interactive variable */
if Q₁ ≠ 0 then
   if Q₂ ≠ 0 then
    if Q₃ ≠ 0 then
       if Q₄≠ 0 then
          (Q₁,Q₂,Q₃,Q₄)
       else
            (Q₁,Q₂,Q₃,α*min(Q₁,Q₂,Q3))
              endif
    else
            if Q₄ ≠ 0 then
               (Q₁,Q₂,α*min(Q₁,Q₂,Q₄),Q4)
        else
            (Q₁,Q₂,α*min(Q₁,Q₂),α*min(Q₁,Q₂))
```

```
          endif
    endif
   else
    if Q₃ ≠ 0 then
        if Q₄≠ 0 then
                    (Q₁,α*min (Q₁,Q₃,Q₄),Q₃,Q₄)
        else
            (Q₁, α*min(Q₁,Q₃),Q₃,α*min (Q₁,Q₃))
               endif
    else
        if Q4≠ 0 then
                    (Q₁,α*min(Q₁,Q₄),α*min (Q₁,Q₄),Q₄)
        else
            (0,0,0,0)
        endif
    endif
   endif
else
  if Q2 ≠ 0 then
   if Q₃ ≠ 0 then
        if Q₄≠ 0 then
            (α*min(Q₂,Q₃,Q₄),Q₂,Q₃,Q₄)
        else
            (α*min(Q₂,Q₃),Q₂,Q₃, α*min(Q₂,Q₃))
        endif
    else
               if Q₄ ≠ 0 then
                (α*min(Q₂,Q₄),Q₂,α*min(Q₂,Q₄),Q₄)
        else
            (α*Q₂,Q₂, α*Q₂,α*Q₂)
        endif
    endif
   else
    if Q₃ ≠ 0 then
               if Q₄≠ 0 then
                (α*min(Q₃.Q₄),α*min (Q₃,Q₄),Q₃,Q₄)
               else
                    (α*Q₃,α*Q₃,Q₃, α*Q₃,α*Q₃)
               endif
    else
        if Q₄≠ 0 then
               (0,0,0,Q₄)
        else
            (0,0,0,0)
        endif
    endif
   endif
endif
```

However, Network Related and Customer Service Quality Parameter result tables of (United) AP state of India only is furnished. Remaining States of India are made available at https://www.dropbox.com/sh/ojwzjgdacwukzon/AADyyYW4zSrSRjMNFF1pwIrwa?dl=0 .

4. RESULTS AND DISCUSSION

The performance of the following methods is compared:

MVG   Moving Average Method[9]
KNN   KNN (nearest-neighbor method)[12]
Mean   Mean imputation [12]
Maximal Maximum relative frequency method (Maximal conditional probability method)[9]
RPQM   Relative Parameter Quantification Method (Proposed Method)

With respect to Table 1, the performance indicators parameters can be as follows.

Rule 1: The values may be minimum for the parameters 3,4,5,8,9,12,13,14,15,16,17
Rule 2: The values may be maximum for the parameters 1, 2,6,7,10,11

4.1 Random values based Test:

Monte Carlo simulation uses random numbers to generate a model. Using these even complex systems can be easily described [20]. All the above methods are compared with RPQM on the hypothetical data set. This technique is called as competitive evolution of models. The goal of this simulation is to choose the best method among above methods. Tables 2, 3 and 4 are used for minimum selection criteria, to take a better business decision. Table 5 and 6 are used for maximum selection criteria, to take a better business decision.

Table 2-A: Random data

Quarter	Player 1	Player 2	Player 3	Player 4	Player 5	Player 6	Player 7	Player 8	Player 9
Q1	1	0.63	0.26	0.01	0.56	0.64	0.2	0.68	
Q2	0.1	0.79				0.62			0.26
Q3	0.3		0.49	0.3	0.18		0.1	0.44	
Q4	0.2		0.34		0.31	0.62			0.7

Table 2-B Missing value estimates									
MVG	0.1	0.71	0.42	0.16	0.25	0.42	0.15	0.56	0.48
KNN	0.1	0.79	0.26	0.01	0.18	0.62	0.1	0.44	0.26
Mean	0.1	0.71	0.36	0.16	0.35	0.63	0.15	0.56	0.48
Maximal	0.1	0.63	0.26	0.01	0.18	0.62	0.1	0.44	0.26
RPQM	0.1	0.47	0.2	0.01	0.14	0.47	0.08	0.33	0.2

Table 2-C Post-imputation estimate of missing value									
MVG	0.1	0.63	0.26	0.01	0.18	0.42	0.1	0.44	0.26
KNN	0.1	0.63	0.26	0.01	0.18	0.62	0.1	0.44	0.26
Mean	0.1	0.63	0.26	0.01	0.18	0.62	0.1	0.44	0.26
Maximal	0.1	0.63	0.26	0.01	0.18	0.62	0.1	0.44	0.26
RPQM	0.1	0.47	0.2	0.01	0.14	0.47	0.08	0.33	0.2

Table 3-A: Random data

Quarter	Player 1	Player 2	Player 3	Player 4	Player 5	Player 6	Player 7	Player 8	Player 9
Q1		0.65		0.21	0.99	0.51	0.75	0.74	0.89
Q2	0.16	0.26	0.35	0.59	0.51	0.45	0.46		
Q3		0.89		0.66		0.5		0.86	0.54
Q4	0.35	0.3	0.37	0.49	0.78	0.7	0.91	0.71	0.7

Table 3-B Missing value estimates									
MVG	0.26	0.26	0.36	0.21	0.59	0.45	0.55	0.79	0.62
KNN	0.16	0.26	0.35	0.21	0.51	0.45	0.46	0.74	0.54
Mean	0.26	0.26	0.36	0.21	0.76	0.45	0.71	0.77	0.71
Maximal	0.16	0.26	0.35	0.21	0.51	0.45	0.46	0.71	0.54
RPQM	0.12	0.26	0.26	0.21	0.38	0.45	0.35	0.53	0.41

Table 3-C Post-imputation estimate of missing value									
MVG	0.16	0.26	0.35	0.21	0.51	0.45	0.46	0.71	0.54
KNN	0.16	0.26	0.35	0.21	0.51	0.45	0.46	0.71	0.54
Mean	0.16	0.26	0.35	0.21	0.51	0.45	0.46	0.71	0.54
Maximal	0.16	0.26	0.35	0.21	0.51	0.45	0.46	0.71	0.54
RPQM	0.12	0.26	0.26	0.21	0.38	0.45	0.35	0.53	0.41

From Tables 2 and 3, it is observed that moving average and KNN method consider fewer values to filling in the missing value. Hence, the value should be always in between minimum and maximum value based on the boundary. Mean imputation method is data dependent and serves good when the missing values are few. Maximal method imputes the existing repeated value so the existed minimum value considered for business decision. It is found that the RPQM method gives or withdraws the support based on the business environment.

Table 4-A: Random data

Quarter	Player 1	Player 2	Player 3	Player 4	Player 5	Player 6	Player 7	Player 8	Player 9
Q1	0.36	0.63	0.21		0.56	0.65	0.2	0.68	
Q2		0.88				0.62			0.26
Q3	0.49		0.51	0.33	0.18		0.1	0.44	
Q4			0.34		0.32	0.62			0.7

4-B Missing value estimates									
MVG	0.43	0.76	0.43	0.33	0.25	0.42	0.15	0.56	0.48
KNN	0.49	0.88	0.51	0.33	0.56	0.62	0.2	0.68	0.26
Mean	0.43	0.76	0.35	0.33	0.35	0.63	0.15	0.56	0.48
Maximal	0.49	0.88	0.51	0.33	0.56	0.62	0.2	0.68	0.7
RPQM	0.37	0.66	0.38	0.33	0.42	0.49	0.15	0.51	0.53

4-C Post-imputation estimate of missing value									
MVG	0.49	0.88	0.51	0.33	0.56	0.65	0.2	0.68	0.7
KNN	0.49	0.88	0.51	0.33	0.56	0.65	0.2	0.68	0.7
Mean	0.49	0.88	0.51	0.33	0.56	0.65	0.2	0.68	0.7
Maximal	0.49	0.88	0.51	0.33	0.56	0.65	0.2	0.68	0.7
RPQM	0.49	0.88	0.51	0.33	0.56	0.65	0.2	0.68	0.7

Table 5-A: Random Data

Quarter	Player 1	Player 2	Player 3	Player 4	Player 5	Player 6	Player 7	Player 8	Player 9
Q1		0.65		0.21	0.99	0.51	0.75	0.74	0.89
Q2	0.16	0.26	0.35	0.59		0.45	0.46		
Q3		0.89		0.66		0.5		0.86	0.54
Q4	0.35	0.3	0.37	0.49	0.78	0.7	0.91	0.71	

5-B Missing Value Estimates									
MVG	0.26	0.89	0.36	0.66	0.89	0.7	0.55	0.79	0.72
KNN	0.16	0.89	0.35	0.66	0.99	0.7	0.91	0.86	0.89
Mean	0.26	0.89	0.36	0.66	0.89	0.7	0.71	0.77	0.72
Maximal	0.35	0.89	0.37	0.66	0.99	0.7	0.91	0.86	0.89
RPQM	0.26	0.89	0.28	0.66	0.74	0.7	0.68	0.65	0.67

5-C Post-imputation estimate of missing value									
MVG	0.35	0.89	0.37	0.66	0.99	0.7	0.91	0.86	0.89
KNN	0.35	0.89	0.37	0.66	0.99	0.7	0.91	0.86	0.89
Mean	0.35	0.89	0.37	0.66	0.99	0.7	0.91	0.86	0.89
Maximal	0.35	0.89	0.37	0.66	0.99	0.7	0.91	0.86	0.89
RPQM	0.35	0.89	0.37	0.66	0.99	0.7	0.91	0.86	0.89

After comparison of hypothetical data for parameter K (K=1, 2, 6, 7, 10, 11) from Tables 4 and 5, the conclusion for Players 1 to 9 is that RPQM and other known methods have been found to be the same i.e. maximum value in the data set.

The method is applied on true data and the following are observed:

[Please refer to the large tables made available for reference:
https://www.dropbox.com/sh/ojwzjgdacwukzon/AADyyYW4zSrSRjMNFF1pwIrwa?dl=0]

The following conclusions have been arrived with the conformity of TRAI performance indicators Table-1. The Network Availability Parameters and the Customer Service Quality parameters of (United) AP state results are as follows.

Parameters 1 and 2 relate to Network Availability, Found that except BSNL, all other service providers satisfied Network Availability.(All values exists for decision taking).

Parameters 3, 4 and 5 relate to Connection Establishment, Shows that all service providers satisfied. (All values exist for decision taking).

Parameters 6, 7 and 8 relate to Connection Maintenance, Shows that BSNL TCH drop (call drop) rate is high. (All values exist for decision taking).

Table 6-A: Performance Values of Metering and billing credibility - postpaid $\leq 0.1\%$ (P-10) maximum required

Qua-rter	Aircel	Airtel	BSNL	Idea	MTS	RCOM CDMA	RCOM GSM	TATA CDMA	TATA GSM	Uni-nor	Video-con	Voda-fone
Q1	0.02	0.00	0.00	0.06	0.00	0.09	0.10	0.00	0.00	0.00	0.00	0.13
Q2	0.01	0.00	0.00	0.06	0.00	0.09	0.10	NR	NR	0.00	0.00	0.05
Q3	0.02	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.05
Q4	0.04	0.00	0.00	0.06	0.00	0.10	0.10	0.00	0.00	NA	NA	0.04

6-B Missing value estimates												
MVG	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13
KNN	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13
Mean	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13
Maximal	0.02	0.00	0.00	0.06	0.00	0.09	0.10	0.00	0.00	0.00	0.00	0.05
RPQM	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13

6-C Post-imputation estimate of missing value												
MVG	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13
KNN	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13
Mean	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13
Maximal	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13
RPQM	0.04	0.01	0.00	0.09	0.00	0.10	0.10	0.00	0.00	0.00	0.00	0.13

Parameter 10 relates to postpaid Metering and Billing. All methods including RPQM show that all service providers satisfied because other quarter values are of the same.

Parameter 11 relates to Prepaid Metering and Billing. All methods including RPQM show that Aircel and Reliance GSM have to improve their performance because other quarter values are of the same.

Parameters 12 relate to Metering and Billing. RPQM show that BSNL, TATA CDMA and TATA GSM have to improve attending and solving billing/charging/validity complaints. Other methods show that all service providers satisfied except BSNL because other quarter values are of it.

Parameters 13 relate to Metering and Billing. As per RPQM TATA CDMA, TATA GSM and Uninor have to improve attending and solving billing/charging/validity complaints. All other methods show that except Uninor all service providers satisfied because other quarter values are of it.

Parameter 14 Accessibility of Call Centre/Customer Care relates to Response Time to Customer for Assistance. RPQM show that TATA CDMA and TATA GSM have to improve their performance. All other methods show that all service providers satisfied.

Parameter 15 relates to Response Time to Customer Assistance. All methods including RPQM show that except Uninor no service providers answered to the customer calls within 60 seconds.

Parameter 16 relates to requests for Termination / Closure of service. RPQM show that Aircel, MTS, TATA CDMA, TATA GSM, Uninor and Videocon have to improve their closing of service. Other methods show that Aircel, TATA GSM, Uninor and Videocon have to improve.

Table 7-A: Performance Values of Time taken for refund of deposits after closures 100% (P-17) Minimum required

Qua-rter	Aircel	Air-tel	BSNL	Idea	MTS	RCOM CDMA	RCOM GSM	TATA CDMA	TATA GSM	Uni-nor	Video-con	Voda-fone
Q1	100	100	100	100	100	100	100	100	100	0	0	100
Q2	100	100	100	67	100	100	100	NR	NR	0	0	100
Q3	100	100	100	100	NR	100	100	99	100	0	0	100
Q4	100	100	100	100	NR	100	100	99	100	0	0	100

7-B Missing value estimates												
MVG	100	100	100	67	100	100	100	99	100	0	0	100
KNN	100	100	100	67	100	100	100	99	100	0	0	100
Mean	100	100	100	67	100	100	100	100	100	0	0	100
Maximal	100	100	100	100	100	100	100	99	100	0	0	100
RPQM	100	100	100	67	75	100	100	74	75	0	0	100

7-C Post-imputation estimate of missing value											
MVG	100	100	100	67	100	100	100	99	100	0	0	100
KNN	100	100	100	67	100	100	100	99	100	0	0	100
Mean	100	100	100	67	100	100	100	99	100	0	0	100
Maximal	100	100	100	67	100	100	100	99	100	0	0	100
RPQM	100	100	100	67	75	100	100	74	75	0	0	100

Parameter 17 relates to Time taken for refund of deposits after closures. RPQM show that Idea, MTS, TATA CDMA, TATA GSM, Uninor and Videocon have to improve the process of refunding deposits after closures. Other methods show that Idea, T-CDMA, Uninor and Videocon have to improve.

Customer satisfaction had been proven lower than expected according to the results acquired by the Network Availability and Customer Service Quality. This implies that telecom service providers in (United) AP India need to improve their efficiency and effectiveness in the provision of telecommunication services that need to meet customer expectation so that customer satisfaction may be enhanced.

From table 2 to 7, based on spot check validation criteria it is observed RPQM performed well than known methods in estimating the missing value whenever their minimum boundary value taken into consideration. Whereas the methods like MVG, KNN, Mean, Maximal and RPQM give similar result when the maximum boundary value is taken into consideration, because the entire methods pick up the maximum value as per the logic and all the results will be similar and a good business decision may not be possible to be taken based on such maximum criteria. This case study work helps the decision makers for developing preferences, minimizes risks and maximizes opportunities with less mathematical effort to make a business decision.

5. CONCLUSION

Missing values in data jeopardize the business decisions and the resulting imputation of missing data is a substantial problem. Inconsistent secondary data sources cause variations in data formats and storage. It is shown how such multiple sources can be used for data analysis through an example. In this paper, using the proposed Relative Parameter Quantification Method to estimate PVs as measures of performance compared with the other existing methods, it is found that the proposed method for the reorganization of data and imputation of data outperformed. This case study concludes this approach to estimate missing values may be adopted for any sort of data with missing values for a better quick business decision process with minimal computational cost.

REFERENCES

[1]   R. Elmasri, S.B. Navathe, Fundamentals of Database Systems. Addison Wesley Pub Co., ISBN 0201542633
[2]   Quantitative Data Cleaning for Large Databases Joseph M. Hellerstein http://db.cs.berkeley.edu/jmh February 27, 2008
[3]   Thomas C. Redman. Data Quality for the Information Age. Artech House, Inc., Norwood, MA, USA, 1996. ISBN 0890068836. Foreword By-A. Blanton Godfrey.

[4]   Thomas C. Redman. The impact of poor data quality on the typical enterprise. Communications of the ACM, 41(2):79{82, 1998. ISSN 0001-0782. doi:http://doi.acm.org/10.1145/269012.269025.

[5]   Taghi Khoshgoftaar, Angela Herzberg, and Naeem Seliya. Resource oriented selection of rule-based classi_cation models: An empirical case study. Software Quality Control,14(4):309{338, 2006. ISSN 0963-9314. doi: http://dx.doi.org/10.1007/s11219-006-0038-1.

[6]   Knowledge creation in marketing: the role of predictive analytics Joe F. Hair JrKennesaw State University, Atlanta, Georgia, USA

[7]   Supply chain management: Concepts and cases Rahul V. Altekar - January 1, 2005 PHI Learning Pvt. Ltd. – Publisher Page No:354-355

[8]   http://www.trai.gov.in/Content/PerformanceIndicatorsReports/1_1_PerformanceIndicators Reports.aspx

[9]   A Comparison of Several Approaches to Missing Attribute Values in Data Mining, Jerzy W. Grzymala-Busse and Ming Hu pp. 378−385, 2001

[10]   "A Statistical Method for Integrated Data Cleaning and Imputation", Chris Mayfield et al, 2009, Purdue e-Pubs. http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2722&context=cstech

[11]   "Estimation of Missing Data Using Computational Intelligence and Decision Trees", George Ssali and Tshilidzi Marwala,2007 http://arxiv.org/ftp/arxiv/papers/0709/0709.1640.pdf

[12]   "Reasoning with Missing Values in Multi Attribute Datasets", Anjana Sharma et al, Vol 3, Issue 5, May 2013, International Journal of Advanced Research in Computer Science and Software Engineering. http://www.ijarcsse.com/docs/papers/Volume_3/5_May2013/V3I5-0219.pdf

[13]   "Some Thought on the MiniMax Principle", Aumann R J et al, January 1972, EBSCOhost Connection http://connection.ebscohost.com/c/articles/7019888/some-thoughts-minimax-principle

[14]   Comparative Decision-Making Analysis- edited by Philip H. Crowley, Thomas R. Zentall-Oxford University Press, 30-Jan-2013 Page no 424-426

[15]   http://www.prenhall.com/divisions/bp/app/russellcd/PROTECT/CHAPTERS/CHAP02S/ CH02SFRM.HTM

[16]   Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis. (2nd ed.). NewYork: Springer-Verlag New York Inc. Section 1.5.

[17]   Lindgren, B.W. (1971). Elements of Decision Theory. New York: The Macmillan Company.

[18]   French, S., & Insua, D.R. (2000). Statistical Decision Theory. London: Arnold.

[19]   http://philosophy.hku.hk/think/strategy/decision.php

[20]   Random Number Generation and Monte Carlo Methods Chapter 2 by James E. Gentle

AUTHORS

Mahesh Kandakatla Profile

Mahesh Kandakatla earned MTech CSE from JNTU Hyderabad; Presently, he is research scholar at OPJS University Rajasthan, India. His active research interests include data mining, network security and adhoc networks.

Prashanth Bolukonda Profile

Prashanth Bolukonda earned MTech CSE from JNTU Hyderabad; Presently, he is Assistant Professor, Vaagdevi College of Engineering, Telangana, India. His active research interests include data mining, network security and adhoc networks.

Dr.Lokanatha C Reddy Profile

Dr. Lokanatha C. Reddy earned M.Sc.(Maths) from Indian Institute of Technology, New Delhi; M.Tech.(CS) with Honours from Indian Statistical Institute, Kolkata; and Ph.D.(CS) from Sri Krishnadevaraya University, Anantapur. Earlier worked at KSRM College of Engineering, Kadapa; Indian Space Research Organization (ISAC) at Bangalore and as the Head of the Computer Centre at the Sri Krishnadevaraya University, Anantapur. Presently, he is a Professor of Computer Science at the Dravidian University, India. His active research interests include Real time Computation, Distributed Computation, Digital Image Processing, Pattern Recognition, Networks, Data Mining, Digital Libraries and Machine Translation.