Paper PO-19

Evaluating effectiveness of management interventions in a hospital using SAS® Text Miner.

Anil Kumar Pantangi, Oklahoma State University, Stillwater, Oklahoma, USA.

Musthan Kader Meeran Mohideen, Oklahoma State University, Stillwater, Oklahoma, USA.

Dr.Goutam Chakraborty, Oklahoma State University, Stillwater, Oklahoma, USA.

Dr.Gary Gaeth, University of Iowa, Stillwater, Iowa, Iowa City, USA.

ABSTRACT

Businesses often implement changes to improve outcomes and enhance customer satisfaction. The best situation occurs when a business can measure the impact of the change before and after the intervention. Healthcare and hospital management is no exception to this. In this paper we have analyzed patient survey data obtained from a large Midwestern University Teaching Hospital. In 2010, the hospital management introduced a key intervention to improve patient satisfaction. We have used SAS Enterprise Guide® and Enterprise Miner® to analyze the pre and post effects of a key intervention – the introduction of an online portal to access patient medical information, including test results. The challenge was to integrate data that were in two forms—quantitative and qualitative (comments). As would often occur, the survey was not specifically designed to measure this intervention. We have analyzed both the aggregate quantitative data and the text data to gauge the specific sources and valence of customers' comments about the intervention. We find a significant increment in the means of outcomes in the aggregate for some variables and we that this appears to be related to when patients commented about test results in the survey This kind of pre and post analyses, where quantitative and qualitative data are used in tandem, can the management in measuring the effectiveness and significance of the intervention strategies.

INTRODUCTION

Interventions are often used by a business to reduce costs, increase efficiency or improve customer's perceptions about their product or services. In this research we evaluate an intervention introduced by a major Midwestern university hospital that was designed to improve access to medical information. In this case, the key to evaluating the value of this intervention is to measure its effectiveness in improving customer satisfaction. The hospital, like many other businesses, routinely surveys its patients using a third-party provider. It is often difficult to change or add questions to such ongoing surveys to specifically capture patient's reactions to specific changes. Prior research has found that analysis of text comments along with quantitative data from a survey increases the explanatory ability of the survey results.[1] Here we illustrate how text mining of the quantitative data can be integrated with the qualitative data (based on patients' comments) to provide deeper insights about changes in customers perceptions pre and post intervention.

Text mining is a complex process because of the unstructured nature of the text data and the complexities associated with handling syntactic and semantic components of text. Text mining studies often include categorization of texts, measuring of valence of texts, predicting texts into predefined categories and information retrieval. [2]. According to Dobson (2010), many companies fail to analyze the patterns in the textual data. However, textual comments and responses provided in these surveys contain a wealth of information.

The hospital investigated in this study has a culture of improvement and lean operational management, so this was one of several efforts to improve performance metrics and enhance patient satisfaction about the visit and clinic. Here we have considered an intervention of 'introduction of an online portal' to view patient medical records, which includes the ability to see results more quickly. The hospital management had gathered quantitative data suggesting that the availability of the online portal has generated a positive impact on patient satisfaction; however, a specific increment in the quantitative survey data was hard to validate because patient satisfaction can be attributed to many factors such as good care by care providers, doctor's treatment or pleasantness in the waiting room and so on.

SURVEY DETAILS

In this study we have used data from a survey sent out by a third party provider to a sample of patients after the completion of an outpatient hospital visit. The survey contains a battery of question in five sections: access to care, during your visit, your care provider, personal issues and overall assessment. Every survey section has quantitative questions based on a 5-point scale and space for handwritten comments, leading to both structured and unstructured data. The overall assessment section has two unstructured questions. Patients are asked to comment about "What was the best thing about your experience with our clinic?" referred to here as "good thing about the clinic" and "Name one thing you wish were different about our clinic" referred to here as the "name a bad thing about the clinic."

For this analysis, we have considered quantitative survey data from one year prior and one year post the intervention. For the unstructured data, we have only considered comment from the "bad thing about the clinic section". The goal of the analysis is to detect whether there are changes in the valence of comments related to obtaining test results in the textual data after the intervention.

METHODOLOGY

QUANTITATIVE SURVEY ANALYSIS

Out of the 35 quantitative questions in the survey, we have analyzed three quantitative questions from several sections of the survey. Of these three variables, O4 and CP10, are considered by the hospital management as key overall metrics on which the management would like to find an impact due to the intervention. The variable I19 is a question that specifically addresses medical test results and therefore is likely to reflect the impact of the intervention. The mean values are computed on a scale of 0 to 100 which are compared for the pre-intervention and post-intervention and reported in Table 1. The sample size of the dataset pre-intervention is 13,302 and the post-intervention dataset is 25,164.

Survey Questions	Variable Name
Likelihood of recommending our Practice to others.	O4
Likelihood of recommending this Care Provider to others	CP10
Access to Test Results	I19

Table 1: Quantitative Survey Questions

Comparison of average values show that for the specific question that assesses access to test results (I19), there is a clear and statistically significant (at 5% level) increase from pre-intervention to post-intervention period. While that is good news for the hospital management, what the management really wants to see is the impact on the overall business metrics. Unfortunately, here the result is murky. In the post-intervention data for the two overall business metrics, one went up significantly (O4) while the other went down significantly (CP10). The actual differences are rather small, but small differences at the high end of the scale are managerially important. However, as with all longitudinal studies, these changes in the overall metrics cannot be purely accounted for by the intervention implemented because many environmental components are changing simultaneously.

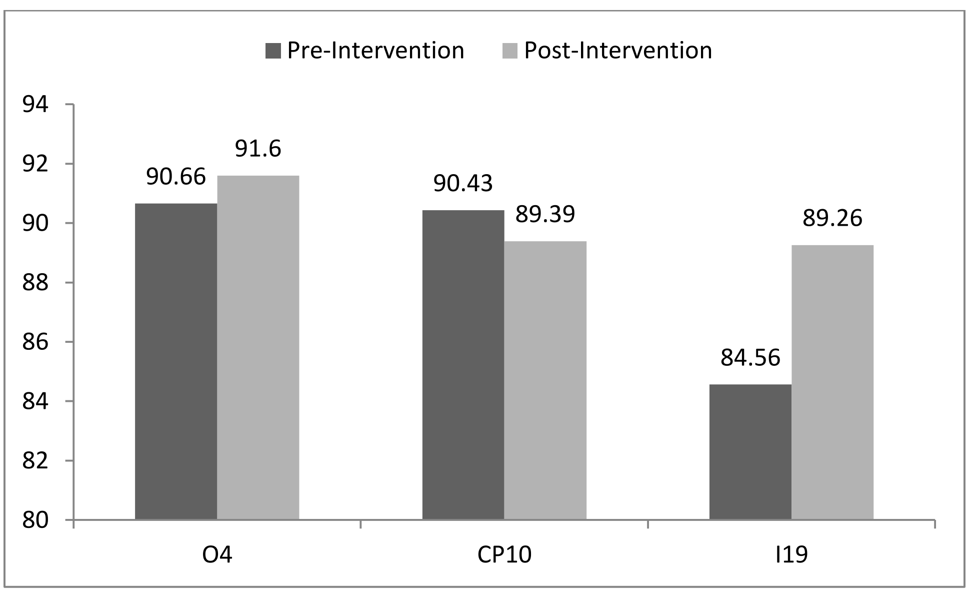
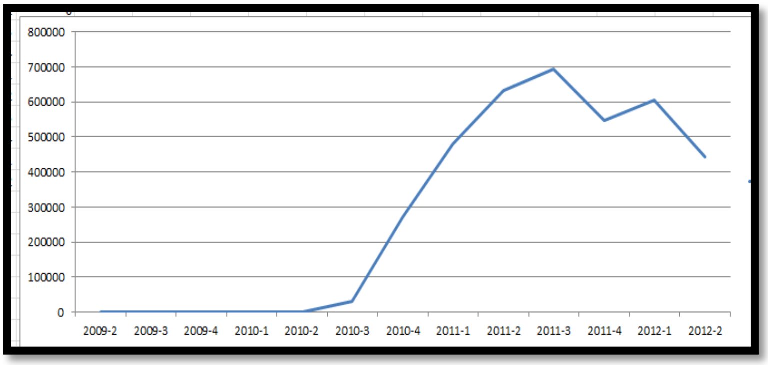


Figure 1: Mean Values Of Business Metrics

To fully explain the effect of the intervention on the quantitative data, the trend of usage of the 'online portal' has been collected and reported in Figure 2. Numbers of instances of the usage of the online portal are captured for each quarter since deployment, and reported below. The graph shows a clear pattern of usage of the portal following the intervention that follows the same trend as the satisfaction in Figures 3 and 4.



Number of instances of access to the portal

Figure 2: Trend of Online Portal Usage.

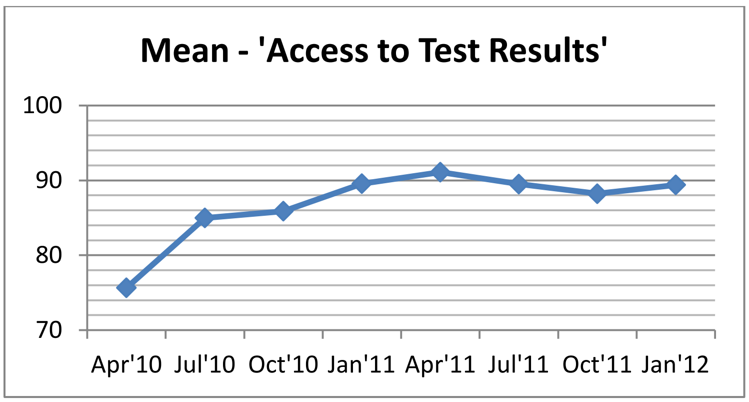


Figure 3: Mean of I19 (Access to Test Results).

It is interesting to find that the average of “access to test results” show an increasing trend (mirroring the pattern in access to portal) starting from first quarter of 2010, right after the intervention.

TOPIC MINING

The quantitative data suggested that access to test result might be a major underlying cause for the overall improvement. To help management better understand these quantitative results, we therefore looked at the negative forms of the comment data to determine if there were fewer negative comments about “test results” after the intervention, thus pointing to this as a potential explanation for the change. The text topic node available in the Text Mining tab of the SAS Enterprise Miner 7.1 is used to extract the themes of relevance from the negative text data. The text topic has to be preceded by the Text parsing and Text filter node as shown in Figure 4 below.

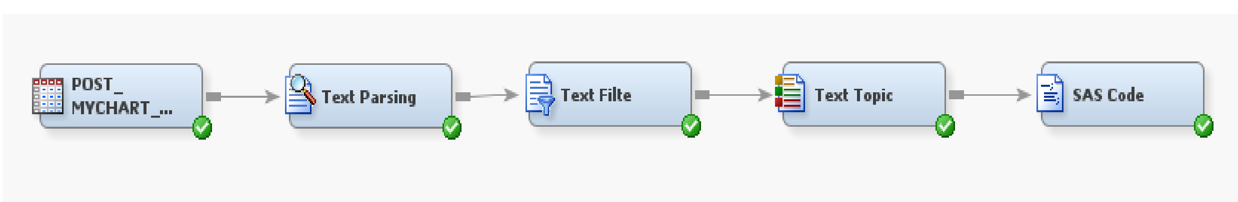


Figure 4: Model Diagram of Topic Mining.

The Text Topic node automatically associates the terms and documents according to both discovered and user-defined topics. Each topic is a collection of terms that pertains to a main theme or idea. Text Topic node assigns a score for each document and terms to each topic. Then, thresholds are used to determine if the association is strong enough to consider whether the document or the terms belong to the topic. As a result, documents and terms may belong to more than one topic or to none at all. [4] Number of topics is typically decided based on the number of documents and terms as well as the business context.

The following settings are used in analyzing text data as shown below in Figure 5.

Property	Value
General	
Node ID	TextParsing3
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Parse	
Parse Variable	comment
Language	English
Detect	
Different Parts of Speech	Yes
Noun Groups	Yes
Multi-word Terms	SASHELP.ENG_MULTI
Find Entities	None
Custom Entities	
Ignore	
Ignore Parts of Speech	'Abbr' 'Aux' 'Conj' 'Det' 'Interj' 'Part' 'Prep' 'Pron
Ignore Types of Entities	
Ignore Types of Attributes	'Num' 'Punct'
Synonyms	
Stem Terms	Yes
Synonyms	SASHELP.ENGSYNMS
Filter	
Start List	
Stop List	SASHELP.ENGSTOP

Property	Value
General	
Node ID	TextFilter3
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Spelling	
Check Spelling	Yes
Dictionary	
Weightings	
Frequency Weighting	Default
Term Weight	Default
Term Filters	
Minimum Number of Documer	4
Maximum Number of Terms	.
Import Synonyms	
Document Filters	
Search Expression	
Subset Documents	
Results	
Filter Viewer	
Spell-Checking Results	EMWS7.TextFilter3_spellDS
Exported Synonyms	

Figure 5: Property Panels of Text Parsing and Text Filter Node respectively.

For the text topic node only the default settings are used. In this analysis we used only multi-term topics. Figure 6 shows the topics discovered by the text topic node. As the default number of topics is 25, many of the topics seem to be redundant. For example, consider the topics 4 & 12 in the pre-intervention comments (left panel in Figure 6) both of which seem to capture the theme of long wait and time spent in the waiting areas.

Identical analysis was conducted for both pre intervention & post intervention comments data in the survey section with 'name a bad thing about the clinic' comments. Topic 22 in the pre-intervention seems to relate to the theme most commonly associated with the introduction of the portal (availability and access of immediate test results).

Topic 23 in the post data analysis relates to negative comments of patients with regards to the availability of their test results.

Topics

Topic ID	Document Cutoff	Term Cutoff	Topic	Category	Number of Terms	# Docs
1	1.014	0.282	closer,home,+home,+wish,+live	Mult	6	325
2	0.812	0.268	+park,+pay,+ramp,+expensive,+fee	Mult	16	412
3	0.792	0.255	+time,+waiting time,+waiting,+short,wait time	Mult	13	359
4	0.795	0.254	+room,+waiting room,+waiting,tv,+big	Mult	13	337
5	0.718	0.252	+exam,+room,+exam room,+room,+cold	Mult	12	273
6	0.677	0.239	+wait,wait time,+time,+short,+time	Mult	18	209
7	0.570	0.230	+appointment,+schedule,+month,+easy,+week	Mult	22	295
8	0.616	0.217	free,+park,+patient,+thing,+wish	Mult	12	319
9	0.540	0.228	+doctor,+wish,+time,different,+doctor	Mult	23	222
10	0.550	0.226	+area,+waiting area,+waiting,+large,+magazine	Mult	19	289
11	0.535	0.216	+wish,dr,closer,different,+rapids	Mult	15	349
12	0.531	0.221	+wait,+long,+month,+wait,+long wait	Mult	24	282
13	0.473	0.218	+clinic,+day,+location,+find,+people	Mult	17	245
14	0.438	0.214	+patient,+schedule,+time,+know,+people	Mult	17	205
15	0.411	0.207	+good,+direction,+magazine,+visit,different	Mult	17	159
16	0.416	0.210	+care,+provider,+care provider,+primary,+problem	Mult	27	158
17	0.422	0.184	+home,closer,+far,illinois,+wish	Mult	6	82
18	0.443	0.201	waiting,time,time waiting,+exam,+long	Mult	19	239
19	0.410	0.205	+hour,+office,+long,+drive,+day	Mult	31	206
20	0.377	0.191	+thing,+people,+day,better,+know	Mult	9	84
21	0.427	0.213	+appt,+schedule,+day,+visit,+wait	Mult	35	368
22	0.393	0.215	+phone,+test,dr,+people,+problem	Mult	53	360
23	0.405	0.191	+live,iowa,+hrs,city,+schedule	Mult	24	100
24	0.384	0.195	+easy,+find,access,+appointment,+hospital	Mult	25	134
25	0.380	0.184	+far,+walk,+wish,+schedule,+drive	Mult	23	123

Topics

Topic ID ▲	Document Cutoff	Term Cutoff	Topic	Category	Number of Terms	# Docs
1	1.147	0.272	+room,+waiting room,+waiting,+magazine,+exam room	Mult	14	818
2	1.193	0.264	closer,+live,+wish,+home,+clinic	Mult	24	973
3	1.131	0.270	+park,+pay,+ramp,better,+cost	Mult	26	1058
4	1.062	0.256	+wait,+wait time,+time,+short,+time	Mult	17	545
5	1.000	0.245	+time,+waiting time,+waiting,+wait time,waiting	Mult	23	979
6	0.892	0.247	+area,+waiting area,+waiting,+big,+large	Mult	20	614
7	0.850	0.219	home,closer,+long,+home,+know	Mult	8	735
8	0.766	0.226	+exam,+exam room,+room,+room,waiting	Mult	19	431
9	0.814	0.213	free,+park,+patient,+long,+easy	Mult	23	716
10	0.692	0.202	+home,closer,+wish,+distance,+mile	Mult	11	447
11	0.700	0.217	+appointment,+schedule,+month,+time,+day	Mult	29	637
12	0.675	0.206	+wish,dr,+different,+clinic,+easy	Mult	15	808
13	0.647	0.214	+clinic,+day,+hospital,+different,+visit	Mult	26	688
14	0.646	0.210	+doctor,+wait,+patient,+know,+time	Mult	32	453
15	0.659	0.210	+wait,dr,+appt,people,+month	Mult	29	650
16	0.653	0.214	+appt,+schedule,+day,+month,+week	Mult	41	710
17	0.572	0.212	+patient,+know,dr,staff,+delay	Mult	47	527
18	0.511	0.190	+thing,+different,+know,+fine,+hospital	Mult	18	235
19	0.525	0.197	+easy,+find,+phone,access,+appt	Mult	32	392
20	0.572	0.199	+hour,+long,+day,+long wait,+wait	Mult	43	613
21	0.526	0.196	+good,+magazine,+communication,+experience,+know	Mult	33	432
22	0.529	0.203	+care,+provider,+care provider,+iowa,+primary	Mult	57	441
23	0.511	0.202	+test,+result,+result,+visit,+test result	Mult	64	732
24	0.442	0.175	+far,+walk,+drive,+live,+travel	Mult	16	175
25	0.469	0.186	waiting,+long,+time,time,+short	Mult	43	554

Figure 6: List of Topics created in Pre (left panel) and Post (right panel) Intervention Text Analysis on the Survey Section – 'Bad thing'.

In order to fully understand the valence of the negativity of the comments, the mean values of the overall business metrics associated with the topics of interest i.e., Topic 22 in the pre-intervention data and Topic 23 in the post-intervention data, are calculated. As can be seen from Figure 7, the averages of both metrics have increased significantly (at 5% level of significance) in the post-intervention. This is quite different from what was found in analyzing the quantitative data alone and clearly shows the positive impact of the intervention on the overall metrics that are key considerations for the hospital management.

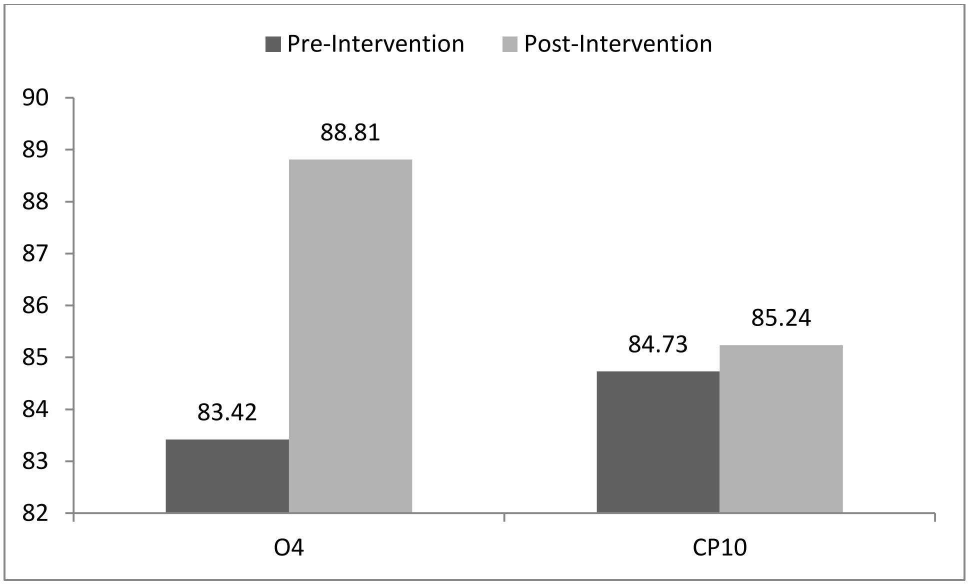


Figure 7: Comparison of Means of Target Variables associated with the relevant topics.

To provide further insights, we find that the percentage of records in the 'name a bad thing about the clinic' section, where patients mentioned about the 'availability of test results' is decreased post the intervention as shown below.

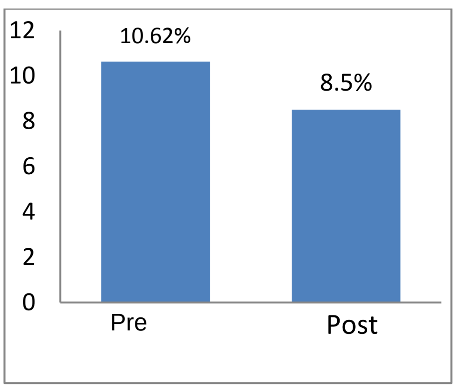


Figure 8: Percentage of records in which 'test results' is mentioned in the 'name a bad thing about the clinic.'

It is also interesting to find that comments about the online portal is generally positive in all of the comments data post intervention as shown below (note that the term is associated with positive sentiments such as easy, love, good, etc.). Overall, it seems the intervention is viewed positively by the patients and it likely contributed to increase in patients' perceptions about the hospital:

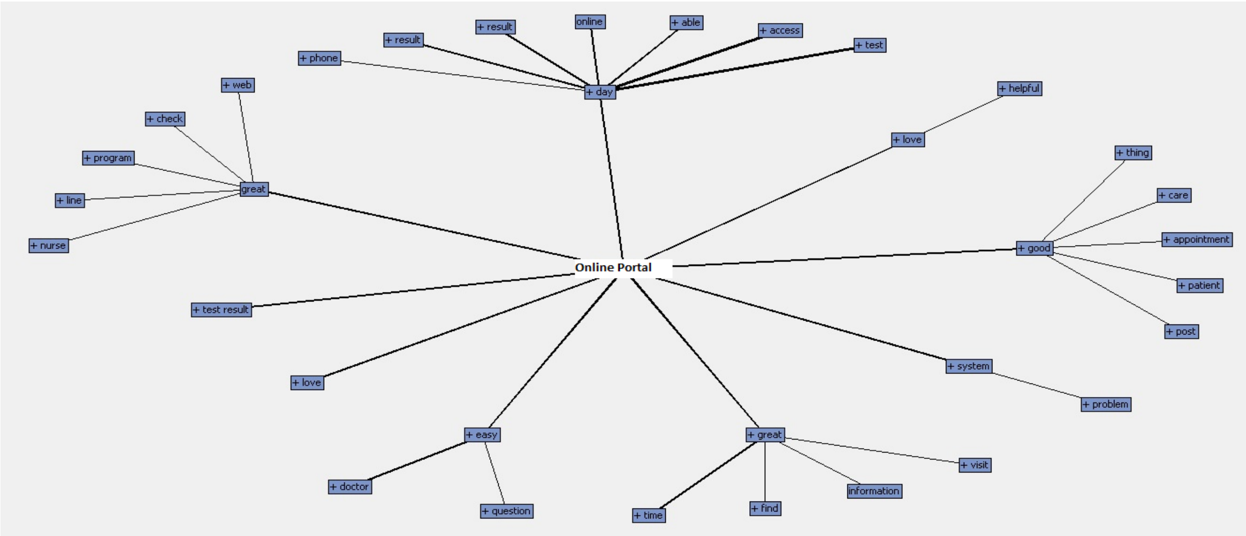


Figure 9: Concept Link Diagram of the term 'Online Portal'.

CONCLUSION

Text mining is a good stand-alone procedure to summarize text and identify the relevant patterns in the comments of any survey. In this research, we demonstrate how deeper insights can be obtained from combining text mining results with quantitative data analysis. In particular, we find that the murky results from the analysis of quantitative data alone improved substantially when text mining results were combined with the numeric data analysis. The text mining reported in this paper mostly used default options in SAS Text Miner and can be improved and fine-tuned by incorporating other options such as including the synonyms or using user-defined topics. In future we are planning to extend this research for the other interventions introduced in the hospital management. We are also planning to use the SAS Sentiment Analysis Studio® to explore the sentiment of the customer perceptions about interventions.

REFERENCES

1. Sullivan, Dan and Ellingsworth, Marty. 2003 "Text Mining Improves Business Intelligence and Predictive Modeling in Insurance." DM Review Magazine.
2. Pantangi and Chakraborty. 2012. "#104-2012 Classification of Customers' Textual Responses via Application of Topic Mining**".** Proceedings SAS® Global Forum 2012.
3. K Markellos, P Markellou, G Mayritsakis. 2009. "Text Mining for Business Intelligence". Available at: http://www.igi-global.com/viewtitlesample.aspx?id=11086
4. Nareddy and Chakraborty. 2011. "#223-2011Improving Customer Loyalty Program through Text Mining of Customers' Comments". Proceedings SAS® Global Forum 2011.
5. Dobson. 2010. David Dobson, Dobson Analytics Inc., Segmenting Textual Data for Automobile Insurance Claims. Proceedings SAS® Global Forum 2010.
6. Miller, W.T. 2005. Data and Text Mining-A Business Applications Approach. Pearson Prentice Hall.

ACKNOWLEDGMENTS

The authors would like to thank the Midwest University Hospital who provided us the data for this research.

RECOMMENDED READING

- Base SAS® Procedures Guide
- SAS® Enterprise Miner Help

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Anil Kumar Pantangi
Enterprise: Oklahoma State University
Work Phone: (203)917-2199
E-mail: anil.pantangi@okstate.edu
City, State ZIP: Stillwater, OK – 74078

Anil Kumar Pantangi is a Masters student at Oklahoma State University. He has worked as Remedy/ITSM Consultant with IBM and Wipro Technologies. He is SAS® Certified Base Programmer for SAS® 9 and SAS® Certified Predictive Modeler.

Name: Musthan Kader Meeran Mohideen
Enterprise: Oklahoma State University
Work Phone: (917)435-9346
E-mail: musthan@okstate.edu
City, State ZIP: Stillwater, OK – 74078

Musthan Kader Meeran Mohideen is a Masters student at Oklahoma State University. He has worked as Product Analyst with Infosys Technologies. He is SAS® Certified Base Programmer for SAS® 9 and SAS® Certified Predictive Modeler.

Name: Dr.Goutam Chakraborty
Enterprise: Oklahoma State University
Work Phone: (405)744-7644
E-mail: goutam.chakraborty@okstate.edu
City, State ZIP: Stillwater, OK – 74078

Goutam Chakraborty is a professor of marketing and founder of SAS and OSU data mining certificate and SAS and OSU business analytics certificate at Oklahoma State University. He has published in many journals such as Journal of Interactive Marketing, Journal of Advertising Research, Journal of Advertising, Journal of Business Research, etc. He has chaired the national conference for direct marketing educators for 2004 and 2005 and co-chaired M2007 data mining conference. He has consulted with companies such as Aetna, Mercruiser, Berendsen Fluid Power, Hilti, Love's travel stop and others. He is also a Business Knowledge Series instructor for SAS.

Name: Dr.Gary Gaeth
Enterprise: University of Iowa
Work Phone: (319) 335-1019
E-mail: gary-gaeth@uiowa.edu
City, State ZIP: Iowa City, IA – 52242

Gary Gaeth is Professor of Marketing and holds the Cedar Rapids Chair of Business. He previously served as Associate Dean of the School of Management for 12 years, where was responsible for the 4 different MBA programs, educating over 1000 students a year. His theory-based research focuses on mathematical models of consumer decision making and his applied research on strategic management in the service businesses. Gary has published widely in marketing, retailing and decision making journals and has consulted with companies in several diverse industries, including Aetna, HNI, American Dental Association, Kirkwood Community College, University of Iowa Hospitals and Clinics, and Orchestra Iowa.