Case Study

SICSR, PUNE

THE BUSINESS VALUE OF TEXT MINING

Abhinav Anil
Prn - 18030142001
Abhishek Anand
Prn - 18030142002
MSc CA(DS)
Sem - III

A Case Study in Text Mining: The Business Value

Abstract

Text mining is an enabling technology that will come to change the process for how businesses derive insights & knowledge from the textual data available to them. The current literature has its focus set on the text mining algorithms and techniques, whereas the practical aspects of text mining are lacking. The efforts of this study aims at helping companies understand what the business value of text mining is with the help of a case study. Subsequently, an SMS-survey method was used to identify additional business areas where text mining could be used to derive business value from. A literature review was conducted to conceptualize the business value of text mining, thus a concept matrix was established. Here a business category and its relative: derived insights & knowledge, domain, and data source are specified. The concept matrix was from then on used to decide when information was of business value, to prove that text mining could be used to derive information of business value.

Text mining analyses was conducted on traffic school data of survey feedback. The results were several patterns, where the business value was derived mainly for the categories of Quality Control & Quality Assurance. After comparing the results of the SMS-survey with the case study empiricism, some difficulties emerged in the categorization of derived information, implying the categories are required to become more specific and distinct. Furthermore, the concept matrix does not comprise all of the business categories that are sure to exist.

Keywords: Text Mining, business value, business value of text mining, survey data analysis

1. Introduction

Studies indicate that 80% of a company's information is contained in text documents. In regards to big data, recent sources give indications of 5% of the data being structured, whereas, 95% is unstructured. Additionally, unstructured data are not only text documents, they are also of formats such as video, image, and audio; therefore, often lacking the traditional structure and organization required by machines for analysis.

Text mining (TM) attempts finding meaningful patterns in unstructured data. The data are usually originating from unstructured text. Other works define it as being focused on finding and extracting meaningful information, knowledge, non-trivial patterns, models, directions, trends or rules from unstructured text documents.

The business value of using text mining (TM) for making sense of data grows apparent when is appearing in larger sizes. Extracting information is harder for humans as the quantity of text grows. Reading only a few sentences or messages out of many for decision-making, may lead to a biased view. The study, therefore, focus on the business value of TM that is derived when the human is no part of the earlier stage, manually making sense of numerous texts. Evidently, the literature present TM as a technology of business value, however, there is little research on the topic conducted with a purpose of demonstrating the business value of TM.

2. Aim and Objectives

Considering the abovementioned, the overall aim of the study is therefore:

"To conduct a purposeful investigation of the business value of TM"

The word purposeful investigation is used as a way to make it clear that the evidence of this study is set to be explicitly oriented toward the business value of TM. The study entails and requires the objective of:

1. To conduct TM analyses, ensuring the business value of TM on a general type of Dataset.

2. To conduct an empirical investigation of companies, identifying general problematic business areas where business value from TM could be derived.

3. To compare the results from the business value as derived from TM the Dataset, with the identified business areas where such business value could be derived.

3. Background

3.1. Text Mining

The subsection of TM, presents some of the more general methods for Text Mining in the literature. When it comes to the subject of Text Mining, much of the literature is too varied to give a clear description of how it is conducted and looks in its actuality The field of Text Mining as will be presented in this section illustrates what TM is capable of by the different methods to attempt clarify how TM could be used for the extraction of information. Knowing how Text Mining can be used in different contexts, will help the reader understand different ways the technology can be of use.

3.1.1 Information Retrieval & Information Extraction

Information Retrieval (IR) could shortly be described as the gathering of, and search for, useful documents in a collection, and the indexing of text. It is an automated process, where all relevant documents are retrieved, simultaneously, mitigating the retrieval of non-relevant documents.

Information Extraction (IE) is a separate method, usually following the use of an IR system. The goal is to transform data from being unstructured to structured, which is more easily digested and analyzed. It either processes unstructured or semi-structured data.

3.1.2 Topic Tracking

Topic tracking systems enables the tracking of documents (or categories) of interest, based on pre-specified or automatically predicted preferences. Topic tracking is applicable in circumstances where companies' wants to be alerted of competitors' or their own activities in news, keep up with competitive products or changes in the market. It can be used as a refinement step, with categorization or text summarization, on a volume of documents, as it could pre-specify the relevance of documents, based on keywords in their content, when searching for information on a topic.

3.1.3 Text Summarization

.

Text summarization is used to convey key information from original text(s) in applications, such as, scientific and news articles, advertisements, emails and blogs. At its core, a summarization has the objectives of determining what the important parts of a text are, followed by, deciding how much of the content is to be reduced. Reducing length and detail, while keeping a document main point, is helpful when the end-user has to quickly judge the document relevancy and worth. A summary can indicate what sources are of relevance, give concise factual information, and give a critical opinion statement on content.

There are two different types of approaches to text summarization. In extractive approaches, a summary is a subset created from the original document. Representatives of sentences, the salient units of a text, are extracted and strung together. Text units are evaluated by analyzing frequency and location in text. It does not require understanding of the text. Abstractive approaches extract semantic information from text. Summaries are of text units not necessarily present in the original text. It is more a complex approach as it incorporates NLP techniques, lexical resources e.g. Word Net, and ontologies, resulting in more coherent summaries.

3.1.4 Categorization

Categorization seeks to identify and classify the main theme of a document by placing said document into a pre-defined set of topics. The numbers of classes are dependent upon on the complexity of taxonomy.

3.1.5 Clustering

Clustering is a technique, which group related documents on the basis of some similarity measure e.g. distance metrics such as k-nearest neighbor (supervised) or k means (unsupervised). The grouping is done automatically, without any pre-specified categorization; thereby differentiating from Categorization. The technique is usually referred to as being unsupervised, but it depends on which specific technique is being adopted. It creates a vector of topics for each document and measures their weights on how well a given document fits into each cluster.

3.1.6 Association Rule Mining

Association rule mining is a technique which attempt finding relationships of variables in a given dataset use an adapted technique of association rule mining, where they measured lifts to find the co-occurrence between different entities, and terms, from forum discussions. It discovers relationships by frequency of two or more recurring entities, in the same sentence or message. It can be used in business to see what items are typically purchased together, and derive a strategy for increased sales, from such information.

3.1.7 Opinion Mining & Sentiment Analysis

Opinion Mining and Sentiment Analysis are techniques which enable the analysis of opinionated text, toward entities such as products, organizations, individuals, and events. The ideal opinion mining tool would: "*process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions about each of them (poor, mixed, good)*".

There are several domains where the technology could be applied. Websites which solicit reviews e.g. Amazon, or IMDB etc. are viable for its use as a source to understand how products or services are perceived. Social media monitoring and analysis could be applied to monitor public relations, gain competitive intelligence, and track company image or product image.

3.2 The Business Value of Text Mining

This section presents the business value of TM as it appears in the literature; revealing different how TM methods and applications, in their relative contexts, lead to a certain business value being derived. First, the notion of the business value of text mining is defined. Then the business value derived from Text Mining in the literature is presented and explained why it is relative to the context of business value of text mining. Subsequently, the concept matrix for the business value of text mining is created from the mini-literature review, and an explanation to how the concept matrix was established is made.

3.2.1 Defining the business value of Text Mining

In the context of management, business value is defined as: "*an informal term that includes all forms of value that determine the health and wellbeing of the firm in the long run*." The use of the term is too ambiguous and vague for a precise use in the context of this study.

The definition of business value in the context of IT, state the business value to be generated from the use of emergent IT-enabled business systems. Emergent IT enabled business systems thus meaning "*informational IT and transactional IT systems and their complementary interactions*". The emergent IT-enabled business system referring to the degree it is: "*able to leverage analytical insights provided by the informational system and embedded in the transactional system*". IT-enabled business systems generate transactional, informational, and strategic benefits.

Transactional benefits include:

- Process efficiencies
- Effectiveness
- Cost reduction

Informational benefits include:

- Fact-based decision making
- Real-time decisions
- Single version of the truth
- Actions based on facts

Strategic benefits include:

- Time to market
- Increased revenue
- Superior customer experience

3.2.2 Quality Control and Quality Assurance

The characteristics of the category "Quality Control and Quality Assurance" is the capability of verifying the quality of a product or service, and how well they currently match a set criteria, or requirements, of the customers, users, subscribers or developers. The benefits are discovering new means of improvements to a product or service, during or after its development in a business setting. A question in accord could be: "*How do we improve a product/service?*"

3.2.3 Customer Relationship

The characteristics of what falls in the category of "Customer Relationship", are the additions to customer service, communications, and customer insights or knowledge. Added benefits are the potential of discovering new means, with an impact on customer satisfaction in a business setting. A question could be: "*How do we increase our customer satisfaction?*"

3.2.4 Comprehensive Summaries of Text

The characteristics of the category "Comprehensive Summaries of Text", is the outsourcing, automation, of activities such as manual analysis of large volumes of texts. The benefits are comprehensive summarized texts, presentation of information, reduced human biases, and reduction of redundant tasks. A question could be: "*What is the general opinion of X-subject?*"

3.2.5 Examples of information derived that are not of business value

This subsection is text mining-derived information, which the author views as examples that does not fit the general notion of a business environment. It serves the purpose of showing contrast, to distinguish, help reduce ambiguity, by demarcating the less relative, when this study refers to “The Business Value of Text Mining”.

3.2.6 A Concept Matrix explaining the Business Value of Text Mining

A systematic approach to reviewing the literature has by identification, selection and extraction of the business value from each article, enabled the formation of three categories: Customer Relationship (CR), Quality Control and Quality Assurance (QC & QA), and Comprehensive Summaries of Text (CST).

4. Method

In the context of a research project define method as the following: “a method refers to an organized approach to problem-solving that includes (1) collecting data, (2) formulating a hypothesis or proposition, (3) testing the hypothesis, (4) interpreting results, and (5) stating conclusions that can later be evaluated independently by others.”

The following chapter has the intention of outlining the chosen research method, and its inherent means. Therefore, answers shall be given to why literature is selected; furthermore, what other forms of data is collected, and description of said data, followed by establishing their purpose and how they are to be used. It is followed by description of how analyses are to be conducted and also the adopted research ethics.

4.1 Text mining case study on dataset of survey feedback

“A method of intensively studying a phenomenon over time within its natural setting in one or a few sites. Especially suitable when there is a desire to understand and explain a phenomenon in a field which is not yet well understood.” It has several methods to data collection, and inferences about the phenomenon of interest tend to be rich, detailed, and contextualized The phenomenon of study in this case study is the: (1) use of TM technologies in a business setting, (2) to derive information of business value.

This case study employ data collection methods such as: secondary data (data collected for other purposes) for drawing inferences with TM technologies. This case study is employed in an interpretive manner for theory building and argues for its theory building as there are no prior similar theories identified, to the knowledge of the author.

Case studies have their inherent weaknesses the authors predict these to be due to heavily contextualized inferences. The latter because the secondary data used in the TM, and the SMS survey, could demonstrate business value for the current organization and context, yet show difficulties in generalizing inferences to other contexts or other organizations. However, this could be established with corroborative case studies. An additional weakness is the replicability of the results.

Considering the Dataset, to replicate the TM analysis results might indicate difficulties in observing the same phenomenon, given the uniqueness and idiosyncrasy of the given case site however, the conclusions of the case research may be possible to replicate. Induction is used to learn from the collected data and build the concepts.

By conducting TM analyses with the use of the R-Language, the investigation has the objective of extracting information of business value from the Dataset, such as the problematized areas on different features or aspects of the given courses. The latter to establish the second dimension, capable of providing evidence in proving the business value of TM. Thereby, the derived information are expected to be means of quality improvement and increased customer satisfaction, for the given service; additionally, a suitable presentation of the information is also of importance. The results of the text mining has the goal of extracting information (of business value) belonging to any of the three TM categories of business value (i.e. CR, QC & QA, and CST).

Its related research question is: "What does TM analyses reveal about the investigated Dataset, and what does the analyses say about the business value of TM, given how well they agree with the concept?"

4.2 SMS-survey method

An empirical investigation of the subscribers is conducted with the use of a SMS survey method. It is a standardized questionnaire to collect data about thoughts and behaviors. According to a survey is suitable if you want to explore perceptions concerning a specific, well-known methodology. The SMS-survey for example attempts two things, (1) identify more categories of business value for TM, (2) the investigation of the methodology regarding deriving value from their information derived by using Analyses SMS.

To clarify, the SMS-survey has the purpose of establishing a third dimension, identifying areas (i.e. categories) in various businesses, to find out where business value from TM could be derived (To avoid confusion, this data is not analyzed with the use of Text Mining). Assumptions are made such that, if the companies collect suitable information via their freeform comment fields, TM could be applied to derive business value from such a source. It adds to the empiricism by finding new business areas where TM could be applied, insofar as the result of the SMS-survey makes it possible. It requires the TM case study to be conducted beforehand, since insights gained from the latter are extrapolated, to infer the applicability of the technology onto the described data.

Its related research question is: "What are the general business areas where TM could be applied to derive business value from?"

4.3 Analysis

The most systematic way of analyzing the collected data is to go through each of the objectives at a time, evaluating said data against responding objective. The proposed research questions are conveniently divisible by their framing, and data are collected with the specific purpose of enabling appropriate answers to each of the said research questions when the goal of the objectives are accomplished.

The analysis is conducted in accordance to a predetermined sequence. (1) First a review of the literature; (2) a TM case study is conducted; (3) an SMS-survey is conducted. Each dimension and their initial conclusions, i.e. the results, are compared and conjoined for drawing the inference, allowing for the argument to the evident business value of TM. An inductive approach is used on the data to make the conclusions. "The process of drawing conclusions based on facts or observed evidence". The intention is to improve the credibility and confidence of the study, by demonstrating triangulation across the collected data.

A model is used to clearly illustrate how the analysis, as a whole, shall be conducted (Observe Figure 1). (1) By using the data with its associated objective, answer each of the initial research questions, (2) derive a result as emerged from the answers, (3) compare the results to answer the main questions, (4) a solution to the fundamental problem is emerged. With the steps (3) and (4), the aim of the study is considered to be achieved.

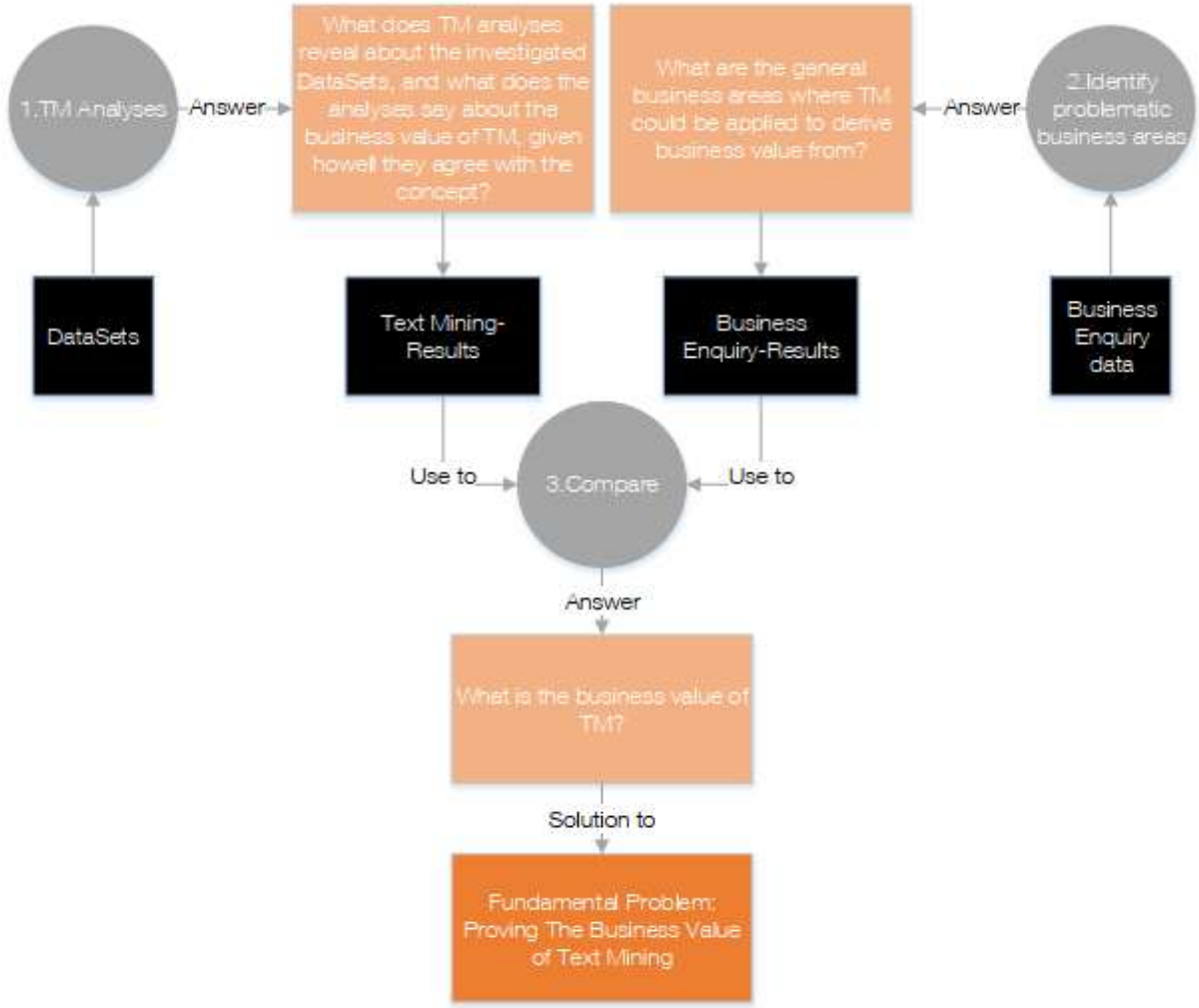

Fig 1 Illustration of the approach for the analysis, in the study.

5. Analysis

The chapter presents the analysis of the conducted investigation, into the business value of TM.

5.1 Text Mining

The subsection is divided into different subsets readable as steps in the method. Each could be viewed as different paradigms to deriving business value with the use of TM.

The word cloud are originally in Swedish, the author translated each word cloud, manually, into English.

5.1.1 Finding patterns by N-grams

Fig 2 Bigram of the DTM

The word clouds for bigrams, (Observe Fig2.) reveal several unique patterns. Some bigrams are reoccurring in a different wording, with similar semantic meaning. Identified patterns are:

- Long time (too much time/time consuming) - 84 Occurrences
- difficult hearing – 17 Occurrences
- bad air – 17 Occurrences
- hot/warm room – 14 Occurrences
- bad ventilation – 11 Occurrences
- cramped room – 11 Occurrences
- cold room – 11 Occurrences

The pattern “long time” is being referred to several times in different bigrams, in different wordings; “long time” is chosen as it is the most frequent.

5.1.2 Sentiment Mining

Sentiment mining was conducted on the Dataset, to see if any additional insight could be derived, with the help of the *tidy text* package. However the results were either equivalent to prior methods or less fruitful. A custom sentiment lexicon for the Swedish language was created from a sentiment lexicon available on the web. The contextual issues for semantic lexicons are apparent, e.g. long is assumed being a positive term, when in fact, in this context it is actually negatively oriented. This could be fixed by the one doing the TM, though the issue of time being a limitation to the study kept this from being addressed. Therefore, given that one would have a more exact sentiment lexicon, it would be viable to follow up on a term, such as "interesting" (Observe Figure 9.), and find its correlations to other words. This would mean a query as such as: What is interesting? And as such the sentiment mining results could be beneficial to the analyses.

The author also attempted translating all of the textual data with the help of the *translate* package through the "Google Translate API 2.0", and employed the same sentiment mining techniques. Here the character translations returned better encoded data, also suitable for bigrams; however, it is uncertain how trustworthy the translations are, and the author concludes the findings as fruitless (Observe Figure 3, Figure 4, Figure 5).

Fig 3 Sentiment mining for positive unigrams On textual data translated to English.

Fig 4 Sentiment mining for negative unigrams On textual data translated to English.

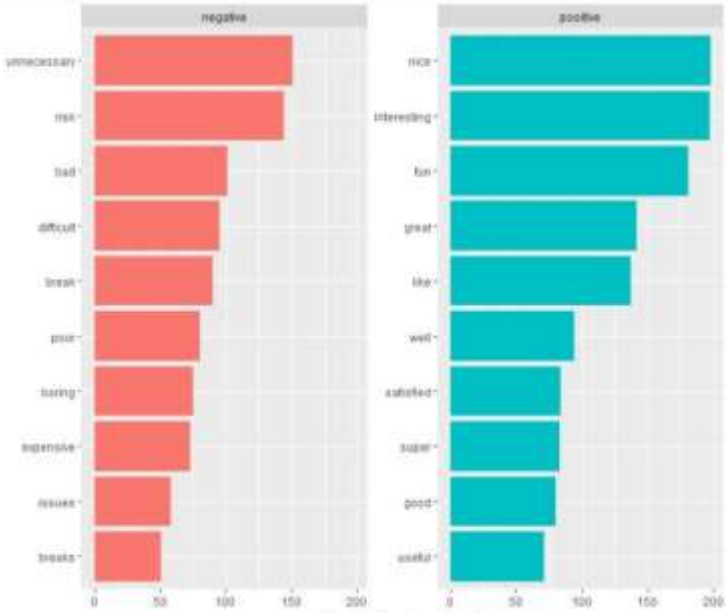

Fig 5 Graphical plot of negative and positive sentiment with *ggplot2* package. They are translated to English.

5.2 SMS-survey

5.2.1 Nominal responses

Feedback/opinions on product	08/05/2017	R1
Feedback/opinions on company	09/05/2017	R2
Feedback/opinions on product	09/05/2017	R3
Customer needs	09/05/2017	R4

Fig. 6. Question (1): What type of information do you gain from the collected free text responses? Identify the type of information that is gained from their respondents (R1-R4 implies which respondent).

Customer Relationship	08/05/2017	R1
Customer Relationship	09/05/2017	R2
Quality Control	09/05/2017	R3
Customer Service	09/05/2017	R4

Fig. 7. Question (2): Does this type of information fall within any general business domains? Specify if this information matches any of the categories of business value of TM or discover additional categories.

By the results of the nominal data, it is easy to conclude that the subscribers use their survey-data results in areas such as Customer Relationship, Customer Service, Quality Control and Quality Assurance. It is clear that the categories of the concept matrix can be improved, since the nominal data give implications to there being an obvious distinction between Customer Service and Customer Relationship. Quality Control and Quality Assurance should similarly be separated as they too are different.

5.2.2 Qualitative responses

The received responses from the qualitative questions asked were few, however, they are still viable for minor analysis. They are translated from Swedish to English.

Question 3. What type of insight is gained by the information you are thinking of? To gain understanding of the information they have in mind, by query of what type of insights are gained from it (i.e. it has the intention of making it clear why the information belongs to a certain category).

"The customers experience of our technicians on location and the process" –Respondent 1
"How to satisfy the demand of the customer" – Respondent 4

The excerpts say the subscribers to Analyze SMS use their collected data to derive insights to improve customer service or customer relationships.

Question 4. "What are the limitations to how you currently derive insights from this information?" Limitations to their current way of analyzing the data.

"It is done manually, but it is not much work. Sadly we don't receive much feedback." -Respondent 1
"The summary could be simplified by being transferred to PowerPoint." –Respondent 4

The first of the two excerpts in regards to the limitations of their data analysis of freeform feedback to have been conducted manually, though because there is so little data being collected it is not so much work.

The second of the excerpts would like the summary of the data to be transferred to PowerPoint; the reason being that analyze SMS strictly presents the data in the web browser, thus makes it difficult to read the entries.

6. Results

The chapter presents the results of the analysis, investigating the business value of TM.

6.1 Information of Business Value

Having conducted text mining on the traffic school data and by using the concept matrix for business value of text mining, several problematic areas in the traffic school courses were revealed. Their issues have been highlighted through discovering patterns in the data, they have been validated with additional techniques such as correlation analysis, and been manually read by searching through the raw data, as a final validation, to be certain that the findings has given actual insights & *knowledge* about these aspects or features of the courses. These insights & knowledge mediate the way toward the improvement of the courses for the traffic school and is therefore the business value that is derived from using text mining.

6.2 Learning from Dataset

It possible conclude that the research objective 1 has been achieved; accomplishing objective 1 we've also learned and answered the related research question; furthermore, the fundamental problem to the study (proving the business value of TM) is solved.

Objective 1. "To conduct TM analyses, ensuring the business value of TM on a general type of Dataset."

Research question. "What does a TM analysis reveal about the investigated Dataset and what does the analyses say about the business value of TM, given how well they agree with the concept?"

As presented in the above subsection, the business value of TM has been given cumulative evidence to its claim. Insights & Knowledge are derived from using TM, and with deductive/inductive reasoning, proven to belong to the categories of business value (as emerged in the literature), showing the concept matrix to be working. The specific business value of each finding can be attributed to some elements found in the concept matrix: *needs*, *wants service/product disapproval*, *technical issues*, and *defect detection*. Additionally it was proven difficult to delineate each of the findings because of their interrelation among some of the categories, in this case Quality Control & Quality Assurance to Customer Relationship. The analysis of the findings categorized each into the category judged to be more appropriate. The reason why they are frequently placed in the category of Quality Control & Quality Assurance, is because the traffic school with certainty do not have recurrent customers, you take the courses, or a course, and do not need to return; therefore, it is arguably so that the information is more oriented toward bettering the experience for future customers, and thus the intention to increase the quality of the service hence them being placed into Quality Control & Quality Assurance. As the data is more oriented or framed toward extracting Quality Control & Quality Assurance related information for the courses, it affects the type of textual data that is collected.

The survey could address this by purposefully querying the customers toward Customer Relationship-related questions; however, the necessity of such a query should be discussed.

6.3 SMS-survey

The SMS-survey had the purpose of accomplishing objective 2 followed by using findings of objective 1 and objective 2, to accomplish objective 3. Additionally the related research question of objective 2 can be answered, but the results are insufficient for drawing valid conclusions.

Objective 2. "To conduct an empirical investigation of companies, identifying general problematic business areas where business value from TM could be derived."

Objective 3. "To compare the results from the business value as derived from TM the Dataset, with the identified business areas where such business value could be derived."

Research question. "What are the general business areas where TM could be applied to derive business value from?

The low responses for the SMS-survey left very little for analysis. What we can learn from the results is that Customer Relationship and Customer Service should be distinct categories, since the former span the entire interactions with the customer, while the latter strictly after purchase. Similarly, Quality Control and Quality Assurance are distinct, since they differ temporally; Quality Control taking part after a product/service has been developed, and Quality Assurance during development. Using the understanding gained by the TM case study, it is possible to conclude the described data to be viable for TM. However, considering one respondent having data where manual analysis is fine, applying TM to the respondent's particular data would be unnecessary.

7. Conclusion

With the emergence of technologies such as text mining, there is a growing importance to understand how the textual data of businesses can be harvested to derive information of business value. The capability of businesses to analyze their large amounts of textual data, not only externally (customers, social media data) but also internally (employee-data), opens up new paths to derive new insight & knowledge. There is little research conducted placed in a business setting that focus on how text mining can be of use for businesses, and most importantly what its part as a constituent component to business operations is or could be. This case study set out to prove the business value of text mining, and made its contribution by using the technology to perform analysis of survey-feedback-data from customers who've partaken in courses given by a traffic school. From this domain, information of business value was derived, specifically the distinct categories of: Quality Control and Customer Relationship. The questions asked were:

- What does a TM analysis reveal about the investigated Dataset and what does the analyses say about the business value of TM, given how well they agree with the concept?
- What are the general business areas where TM could be applied to derive business value from?

An SMS-survey was conducted though the results were not as desirable. The response frequency was too low, and limited the ability to draw any conclusion of substance. However, by its design the conclusion is made that the categories to the concept matrix should become more explanatory by making distinction between terms such as Quality Control & Quality Assurance, and Customer Relationship & Customer Service.

8. References

Abdous, M'hammed., & He, Wu. (2011). Using text mining to uncover students' technology-related problems in live video streaming. British Journal of Educational Technology. 42 (1). 2011. Pp. 40-49.

Abrahams, A. S., Fan, W., Alan Wang, G., Zhang, Z. J., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. Production and Operations Management, 24 (6). June 2015. Pp. 975-990.

Berndtsson, Mikael., Hansson, Jörgen., Olsson, B., Lundell, Björn. (2008). Thesis Projects : A Guide for Students in Computer Science and Information Systems. 2nd ed. Springer.

Bhattacherje, A. (2012). Social Science Research: Principles, Methods and Practices: Creative Commons Licence.

Choi, D., Ko, B., Kim, H., & Kim, P. (2013). Text analysis for detecting terrorismrelated articles on the web. Journal of Network and Computer Applications, 38. Pp.16-21.

Cukier, K., (2010). Data, data everywhere: A special report on managing information. The Economist. February 27.

Dave, Kushal., Lawrence, Steve., & Pennock, M, David. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. WWW '03 Proceedings of the 12th international conference on World Wide Web. May, 2003. Pp. 519-528.

Feinerer, Ingo., Hornik, Kurt., & Meyer, David. (2008). Text Mining Infrastructure in R. Journal of Statistical Software, 25 (5). March. 2008.

Feldman, Ronen., & Dagan, Ido. (1995). Knowledge Discovery in Textual Databases (KDT). KDD-95 Proceedings. 1995.

Fuller, Christie, M., Biros, David, P., and Delen, Dursun. (2011). An investigation of data and text mining methods for real world deception detection. Expert Systems with Applications, 38. 2011. Pp. 8392-8398.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35. Pp.137-144.

Rose, J & Lennerholt, C. (2017). Low Cost Text Mining as a Strategy for Qualitative Researchers. The Electronic Journal of Business Research Methods. Volume 15 Issue 1 2017. Pp. 48-62.

Someh, Asadi, Ida & Shanks, Graeme. (2016). Emergence of Business Value from Complementary Interactions between Informational and Transactional IT systems. Australasian Journal of Information Systems. Volume 20, Selected Papers from the ISF Conference.

Xiang, Z., Schwarts, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? International Journal of Hospitality Management, 44. Pp. 120-130.