

## The Application of Text Mining in Business Research

Björn Preuss\*

Department of International Economics and Management, Copenhagen Business School, Denmark

### Abstract

The aim of this paper is to present a methodological concept in business research that has the potential to become one of the most powerful methods in the upcoming years when it comes to research qualitative phenomena in business and society. It presents a selection of algorithms as well elaborates on potential use cases for a text mining based approach to qualitative data analysis.

**Keywords:** Business; Text mining; Marketing

### Introduction

In recent years, the amount of qualitative data growth exponentially. E-mail, posts, tweets and all another sort of textual data gets created every day. To analyze this amount of data gets more and more interesting for research but also for business. Especially in marketing but also leadership and the behavioral sciences is the concept of text mining evolving.

In this paper, the researcher aims to give an indication of what is nowadays possible and how textual data could be brought into a useful form both for business as well as for research. One recent application of this type of concept is digital marketing. By using text analytics and in specific support vector machines, the marketer can get an overview of sentiments and know how potential leads reaction on an online marketing campaign by analyzing the comments on the posts [1]. The core concept of text mining is based on counting words in a given text corpus. The occurrence of words and combination of words gets in the following analyzed by using certain algorithms [2]. This method of text mining is an interdisciplinary field and combines techniques from statistics, linguistics, data mining and information technology [2,3]. The following paper will discuss certain methods that could be used to trace information from a textual corpus and it will sort these techniques and algorithms into groups.

### Discussion of Method

The base is always the text itself and the techniques used to extract the information out of the text. The reason why this is necessary is because the vast amount of textual data makes it impossible for a researcher or manager to read all this information [2]. And even if this would be possible, it would take so long that the information from the first text would get lost on the way. Because of this, the information that is contained in the text will be extracted and put into a form that can be analyzed by machines.

The first step that needs to be done for this is tokenization. That means that the text is split into single words. At this stage, there are different possibilities that can be done based on the further analysis. The words can be combined to n-grams that mean that words which are used together will be linked since they might contain a certain sense [4]. They can also be traced down to their word-stem. This stemming makes it possible to combine sentences in different tenses without having them in different grammatical forms [5]. Another step that usually needs to be taken is that stop and small words need to be deleted since they do not contain relevant information of the text and do not need to be analyzed by the machine. The result of this text

mining steps is a table of words that summarize the core information of the text (Table 1).

### How to Make Sense of the Words

After splitting the text up into the words, this list of words or n-grams needs to be analyzed to filter out the main meaning of the text. Therefore, different machine learning and clustering algorithms can be used. Generally, the algorithms can be split into supervised and unsupervised learning. Supervised means that the algorithm gets first trained on a pre-labeled set of data [4]. The goal is then to sort the rest of the data set into the predefined clusters by finding similarities in the data set [6,7]. These sort of algorithms can also be applied to different kind of datasets, however, in this case, we want to focus on the application with textual data. Unsupervised means in this case that there are not predefined clusters and the algorithms sort the data into clusters by which each cluster contains the samples that show the greatest similarities. This family of algorithms gets more used when it comes to an explorative approach. This could be for example if a company collects information about the leadership style and the main aspects of it should be analyzed. The algorithm would then present the groups of textual statements that are most similar. With this approach, the management would know what the respondents (those who give the statements) think about the leadership. A list of groups of algorithms can be found in the appendix of this paper.

One application of text mining in research is the field of leadership and corporate culture. Since this field is really soft and not well measurable, text mining creates a great opportunity to analyze qualitative statements about the leadership style within an organization [8]. Traditionally cultural research focus on countries and uses questionnaires to get the relevant data [9,10]. The need for a flexible and company centered approach leads to the search for new methods in this research discipline. The author of this appears applied this method recently on the data of 50 organizations to analyze the impact that different leadership approaches have on M&A transactions. Here for, qualitative statements from employees were collected in which

\*Corresponding author: Björn Preuss, Department of International Economics and Management, Copenhagen Business School, Denmark, Tel: +46 7683 49492; E-mail: [bp.int@cbs.dk](mailto:bp.int@cbs.dk)

Received April 27, 2017; Accepted May 02, 2017; Published May 09, 2017

Citation: Preuss B (2017) The Application of Text Mining in Business Research. J Account Mark 6: 232. doi: [10.4172/2168-9601.1000232](https://doi.org/10.4172/2168-9601.1000232)

Copyright: © 2017 Preuss B. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Name	Group	Type	Aim
k-Means, fuzzy-c-means	Clustering	Unsupervised	Discovering structure
One-class-SVM, PCA	Anomaly detection	Unsupervised	Finding unusual data points
Mult. Linear regression, Mult. A neural network, Mult. Decision jungle, Mult decision tree, One-v-all multiclass	Multi-class classification	Supervised	Predicting multiple categories
Two-class-SVM, Two-class average, perception, Two class log. Regression, Two-class Bayes classifier	Two class classification	Supervised	Predicting two categories
Various models	Regression	Regression	Predicting values

Table 1: Types of machine learning algorithms.

they describe the working place and the management style. This textual data which contained ca. 40 statements per company got analyzed using text mining. Since the literature already stated an impact of corporate culture on M&A transactions, the first goal of this analysis was to support this hypothesis by using the text mining method. A second question that this paper addressed, has been in which specific way corporate culture impacts M&A transactions. Previous studies addressed only the question if there is an impact but not which cultural phenomena would be enhancing for M&A transactions. Having this in mind, a non-supervised clustering algorithm was used to figure out how culture in general looks in companies that are more successful when it comes to M&A transactions.

The findings of this step needed to be tested on whether there is a significant relationship between this defined role models of culture named the M&A enhancing and the non-enhancing one. Before testing this relationship in a multivariate regression model, a supervised learning classifier was used to label the whole dataset based on to which group of culture a statement belongs. The result of this project was that at first, corporate culture had a significant impact on the creation of successful M&A transactions. Furthermore, the role models produced regression models with R<sup>2</sup> of over 40% [11].

### The Biggest Advantage of Text Mining

The biggest advantage of text mining is that it can be used to analyze actual existing data without designing a special survey that needs to be sent out and where a min number of responses needs to be collected. In contrast, most textual data in organizations but also in our daily life gets produced with a different purpose. This means that it reflects much more the reality and is less biased than a specially designed survey. Surveys are generally quite expensive and specially designed for one purpose. Furthermore, the respondent gets influenced by the person who asks the questionnaire and does not necessarily know directly what to answer. Many studies rely therefore on low data quality which makes the studies less reliable. Now when it comes to the text mining approach, no new data needs to be produced. The method aims to analyze the existing data that was produced along another purpose in the real world. This means that analyzing this corpus of real-world produced data would present a much more realistic picture of the reality since it is not artificially produced just for the research purpose. With this, text analysis takes qualitative research one step closer to the sort of research that is done till now in data-heavy disciplines like finance but also in other disciplines where measurement of phenomena is usual.

This approach creates, in addition, the possibility to run such analysis on a regular base. Since the data does not need to be collected and the analysis steps can be automated, it is possible to analyze changes

in the perceived corporate culture or leadership style. This might help researchers but also decision makers to better understand this complex field of leadership and management. Having this two main arguments in mind, there can be seen more and more applications of text analytics in the business field.

### References

- Friedrich N (2015) Applying sentiment analysis for tweets linking to scientific papers. GRIN Verlag.
- Feldman R, Sanger J (2007) The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.
- Hippner H, Rentzmann R (2006) Text Mining. Informatik-Spektrum 29: 287-290.
- Ignatow G, Mihalcea R (2016) Text Mining: A Guidebook for the Social Sciences. SAGE Publications.
- Järvelin A, Järvelin A, Järvelin K (2007) s-grams: Defining generalized n-grams for information retrieval. Information Processing & Management 43: 1005-1019.
- Dumais S, Platt J, Heckerman D, Sahami M (1998) Inductive learning algorithms and representations for text categorization. In: Proceedings of the seventh international conference on Information and knowledge management-CIKM '98.
- Yang Y, Liu X (1999) A re-examination of text categorization methods. In: Proceedings of the 22<sup>nd</sup> annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99.
- Miner G (2012) Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications. Academic Press.
- Chhokar JS, Brodbeck FC, House RJ (2013) Culture and Leadership Across the World: The GLOBE Book of In-Depth Studies of 25 Societies. Routledge.
- Hofstede G (2011) Dimensionalizing Cultures: The Hofstede Model in Context. Online Readings in Psychology and Culture.
- Li F, Villani M (2013) Efficient Bayesian Multivariate Surface Regression. Scandinavian Journal of Statistics, Theory and Applications 40: 706-723.

### OMICS International: Open Access Publication Benefits & Features

**Unique features:**

- Increased global visibility of articles through worldwide distribution and indexing
- Showcasing recent research output in a timely and updated manner
- Special issues on the current trends of scientific research

**Special features:**

- 700+ Open Access Journals
- 50,000+ editorial team
- Rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at major indexing services
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: <http://www.omicsonline.org/submission>

Citation: Preuss B (2017) The Application of Text Mining in Business Research. J Account Mark 6: 232. doi: 10.4172/2168-9601.1000232