# Text Analysis Applications:

1. Risk management: No matter the industry, Insufficient risk analysis is often a leading cause of failure. This is especially true in the financial industry where adoption of Risk Management Software based on text mining technology can dramatically increase the ability to mitigate risk, enabling complete management of thousands of sources and petabytes of text documents, and providing the ability to link together information and be able to access the right information at the right time.
2. Knowledge management: Not being able to find important information quickly is always a challenge when managing large volumes of text documents—just ask anyone in the healthcare industry. Here, organizations are challenged with a tremendous amount of information—decades of research in genomics and molecular techniques, for example, as well as volumes of clinical patient data—that could potentially be useful for their largest profit center: new product development. Here, knowledge management software based on text mining offer a clear and reliable solution for the “info-glut” problem.
3. Cybercrime prevention: The anonymous nature of the internet and the many communication features operated through it contribute to the increased risk of internet-based crimes. Today, text mining intelligence and anti-crime applications are making internet crime prevention easier for any enterprise and law enforcement or intelligence agencies.
4. Customer care service: This is one of the most “traditional” applications of text mining technology. Today, text analytics software is frequently adopted to improve customer experience using different sources of valuable information such as surveys, trouble tickets, and customer call notes to improve the quality, effectiveness and speed in resolving problems. Text analysis is used to provide a rapid, automated response to the customer, dramatically reducing their reliance on call center operators to solve problems.
5. Fraud detection through claims investigation: Text analytics is a tremendously effective technology in any domain where the majority of information is collected as text. Insurance companies are taking advantage of text mining technologies by combining the results of text analysis with structured data to prevent frauds and swiftly process claims.
6. Contextual Advertising: Digital advertising is a moderately new and growing field of application for text analytics. Here, companies such as Admantx have made text mining the core engine for contextual retargeting with great success. Compared to the traditional cookie-based approach, contextual advertising provides better accuracy, completely preserves the user’s privacy.
7. Business intelligence: This process is used by large companies to uphold and support decision making. Here, text mining really makes the difference, enabling the analyst to quickly jump at the answer even when analyzing petabytes of internal and open source data. Applications such as the Cogito Intelligence Platform (link to CIP) are able to monitor thousands of sources and analyze large data volumes to extract from them only the relevant content.
8. Content enrichment: While it’s true that working with text content still requires a bit of human effort, text analytics techniques make a significant difference when it comes to being able to more effectively manage large volumes of information. Text mining techniques enrich content, providing a scalable layer to tag, organize and summarize the available content that makes it suitable for a variety of purposes.
9. Spam filtering; E-mail is an effective, fast and reasonably cheap way to communicate, but it comes with a dark side: spam. Today, spam is a major issue for internet service providers, increasing their costs for service management and hardware\software updating; for users, spam is an entry point for viruses and impacts productivity. Text mining techniques can be implemented to improve the effectiveness of statistical-based filtering methods.
10. Social media data analysis: Today, social media is one of the most prolific sources of unstructured data; organizations have taken notice. Social media is increasingly being recognized as a valuable source of market and customer intelligence, and companies are using it to analyze or predict customer needs and understand the perception of their brand. In both needs Text analytics can address both by analyzing large volumes of unstructured data, extracting opinions, emotions and sentiment and their relations with brands and products.
11. Enterprise Business Intelligence/Data Mining, Competitive Intelligence
12. E-Discovery, Records Management
13. National Security/Intelligence
14. Scientific discovery, especially Life Sciences
15. Sentiment Analysis Tools, Listening Platforms
16. Natural Language/Semantic Toolkit or Service
17. Publishing
18. Automated ad placement
19. Search/Information Access
20. Social media monitoring
21. Security applications: Many text mining software packages are marketed for security applications, especially monitoring and analysis of online plain text sources such as Internet news, blogs, etc. for national security purposes.It is also involved in the study of text encryption/decryption.
22. Biomedical applications: Main article: Biomedical text mining - A range of text mining applications in the biomedical literature has been described.One online text mining application in the biomedical literature is PubGene that combines biomedical text mining with network visualization as an Internet service. GoPubMed is a knowledge-based search engine for biomedical texts.
23. Software applications: Text mining methods and software is also being researched and developed by major firms, including IBM and Microsoft, to further automate the mining and analysis processes, and by different firms working in the area of search and indexing in general as a way to improve their results. Within public sector much effort has been concentrated on creating software for tracking and monitoring terrorist activities.
24. Online media applications: Text mining is being used by large media companies, such as the Tribune Company, to clarify information and to provide readers with greater search experiences, which in turn increases site "stickiness" and revenue. Additionally, on the back end, editors are benefiting by being able to share, associate and package news across properties, significantly increasing opportunities to monetize content.
25. Marketing applications: Text mining is starting to be used in marketing as well, more specifically in analytical customer relationship management. Coussement and Van den Poel apply it to improve predictive analytics models for customer churn (customer attrition).
26. Sentiment analysis: Sentiment analysis may involve analysis of movie reviews for estimating how favorable a review is for a movie. Such an analysis may need a labeled data set or labeling of the affectivity of words. Resources for affectivity of words and concepts have been made for WordNet and ConceptNet, respectively. Text has been used to detect emotions in the related area of affective computing. Text based approaches to affective computing have been used on multiple corpora such as students evaluations, children stories and news stories.
27. Academic applications: The issue of text mining is of importance to publishers who hold large databases of information needing indexing for retrieval. This is especially true in scientific disciplines, in which highly specific information is often contained within written text. Therefore, initiatives have been taken such as Nature's proposal for an Open Text Mining Interface (OTMI) and the National Institutes of Health's common Journal Publishing Document Type Definition (DTD) that would provide semantic cues to machines to answer specific queries contained within text without removing publisher barriers to public access.

Academic institutions have also become involved in the text mining initiative:

The National Centre for Text Mining (NaCTeM), is the first publicly funded text mining centre in the world. NaCTeM is operated by the University of Manchester[23] in close collaboration with the Tsujii Lab,[24] University of Tokyo.[25] NaCTeM provides customised tools, research facilities and offers advice to the academic community. They are funded by the Joint Information Systems Committee (JISC) and two of the UK Research Councils (EPSRC & BBSRC). With an initial focus on text mining in the biological and biomedical sciences, research has since expanded into the areas of social sciences.

In the United States, the School of Information at University of California, Berkeley is developing a program called BioText to assist biology researchers in text mining and analysis.

1. Digital humanities and computational sociology: The automatic analysis of vast textual corpora has created the possibility for scholars to analyse millions of documents in multiple languages with very limited manual intervention. Key enabling technologies have been parsing, machine translation, topic categorization, and machine learning. The automatic parsing of textual corpora has enabled the extraction of actors and their relational networks on a vast scale, turning textual data into network data. The resulting networks, which can contain thousands of nodes, are then analysed by using tools from network theory to identify the key actors, the key communities or parties, and general properties such as robustness or structural stability of the overall network, or centrality of certain nodes.[27] This automates the approach introduced by quantitative narrative analysis,[28] whereby subject-verb-object triplets are identified with pairs of actors linked by an action, or pairs formed by actor-object.

Content analysis has been a traditional part of social sciences and media studies for a long time. The automation of content analysis has allowed a "big data" revolution to take place in that field, with studies in social media and newspaper content that include millions of news items. Gender bias, readability, content similarity, reader preferences, and even mood have been analyzed based on text mining methods over millions of documents.[29][30][31][32] The analysis of readability, gender bias and topic bias was demonstrated in Flaounas et al.[33] showing how different topics have different gender biases and levels of readability; the possibility to detect mood shifts in a vast population by analysing Twitter content was demonstrated as well.

1. Narrative Network of Elections
2. Spam filtering
3. Creating suggestion and recommendations (like amazon)
4. Monitoring public opinions (for example in blogs or review sites)
5. Customer service, email support
6. Automatic labeling of documents in business libraries
7. Measuring customer preferences by analyzing qualitative interviews
8. Fraud detection by investigating notification of claims
9. Fighting cyberbullying or cybercrime in IM and IRC chat
10. Text Mining in the Automotive Industry: It’s been estimated that warranties cost automotive companies more than $35 billion in the U.S. annually. Considering this tough environment, it is imperative that auto companies explore all opportunities for reducing costs. Optimizing warranty cost is a very important lever in the cost equation for automobile manufacturers. If one is able to get even a marginal improvement in money spent in warranty cost, it can have a multiplier effect on the overall bottom line. One of the most underutilized dimensions of optimizing warranty cost is input from service technicians’ comments. From those comments, the text mining process can surface nuggets of component defect insights yielding interventions for preventing them in future.
11. Five Possible Actions to Trigger after Text Mining Technicians’ Comments

Once we know the answers to these questions through a structured text mining process, automobile companies can take four follow-up actions which will reduce warranty-related cost erosion, optimize dealer inventory for spare parts and help suppliers deliver quality components:

Auto component sourcing decisions: Auto manufacturers can share the results of text mining technicians’ comments with specific product suppliers and undertake joint initiatives to reduce the number of defective components.

Early warning system: Automobile companies can consider building an early warning system based on frequency of occurrence of specific keywords in a watch list like “brake lining,” ”short circuit,” etc. which could prove fatal and cause legal liability in some cases.

Optimize internal manufacturing processes: If the component in question was manufactured internally, then the specific manufacturing process responsible for the defective component can be re-examined/re-engineered to eliminate reoccurrence.

Inventory optimization: The frequency of occurrence of select spare parts/auto components can be used as an input to forecast the regional need for spare parts (auto spare part inventory optimization).

Defect taxonomy structure modifications: Most dealer management systems have a preliminary taxonomy to classify defects. This taxonomy may need changes to it depending upon how well it was originally defined. If most defects get classified under miscellaneous, then keyword and theme frequency analysis after text mining can guide in the creation of new defect classifications.

1. Text Mining in the Healthcare Industry: Most countries typically spend anywhere between 3-10% of their GDP on healthcare. The healthcare industry is a huge spender on technology and, with the proliferation of hospital management systems and low-cost devices to log patient statistics, there is a sudden increase in the breadth and depth of patient data. By mining the comments of doctors’ diagnosis transcripts, outputs can yield information that benefits the healthcare industry in numerous ways, such as:

Isolating the top 10 diseases by keyword frequencies per region and leveraging the findings to optimize the mix of tablets/medicines to stock on the limited outlet shelf, keeping in mind the changes in frequency of disease related keywords.

Based on doctors’ comments, an early warning system can be woven within text mining outputs to detect sudden changes to “chatter” from doctors regarding specific diseases. For example, if the frequency of the keyword lungs or breathing exceeds 45 appearances in the last 30 days for a given ZIP code or region, it can be a clue to excessive environmental conditions which are resulting in respiratory problems. A proactive intervention can be activated to remedy the situation.Dasfdfadfsd

1. Text Mining in the Credit Card Industry: With the proliferation of credit cards, companies need to do the difficult balancing act of identifying which card features (i.e., line of credit, billing cycle, outlet points and coverage) are resonating with customers and, at the same time, minimize the number of defaults/recovery related interventions. Text mining can help optimize both the collection process as well as the customer experience optimization process.

A top ten complaint keyword watch list can be generated by mining the inbound customer service rep (CSR) call transcripts on a daily basis. From this, you can filter out keywords that were expressed by high-value customers. For example, if the keyword billing error occurs for customers with a credit limit over $200,000, then relationship managers can call the customer and put interventions into the billing process to help prevent reoccurence.

Text mining can also be used to rate call center staff performance. As an example, a large credit card company in the U.S. had about 600 call center reps receiving inbound calls. Every rep was expected to enter verbose comments to record the nature of the call, but not all were entering detailed text. On one end of the spectrum, there were call center representative entering an average 5 to 6 lines, whereas on the other hand, there were a few who entered just 3 to 5 words. As a result, the organization was missing out on valuable intelligence if only sparse text was recorded. A text mining process was built which gave keyword frequency count by call center representatives. The bottom decile had to undergo additional training to ensure that they entered detailed text, which is valuable for the credit card company.

1. Publishing and media.
2. Telecommunications, energy and other services industries.
3. Information technology sector and Internet.
4. Banks, insurance and financial markets.
5. Political institutions, political analysts, public administration and legal documents.
6. Pharmaceutical and research companies and healthcare.
7. Competitive Intelligence: Organizations today are very keen to know about their performance in the market with respect to the products and services they offer to its customers. They want to collect information about themselves in order to find out if there is any need to reorganize and restructure their strategies according to market demands and also to the opportunities that the market presents. In addition to this they are also interested in collecting the information about the market and their competitors. They also have to manage huge collection of data, process and analyze this data to get useful insights and make new plans. The goal of Competitive Intelligence is to extract only relevant information from various relevant data sources. Once the material is collected, it is classified into categories to develop a database, and analyzing the database to get answers to specific and crucial information for company strategies. The typical queries concern the products, the sectors of investment of the competitors, the partnerships existing in markets, the relevant financial indicators, and the names of the employees of a company with a certain profile of competencies. Organizations, prior to having a text mining system, would have a department that would dedicatedly look into the continuous monitoring of information (financial, geopolitical, technical and economic) and answer the queries coming from the different business areas by the use of manual operation. The process of manually compiling documents according to a user's needs and preferences and into actionable reports is very labor intensive, and is greatly amplified when it needs to be updated frequently. With the introduction of text mining systems the return on investment was evident when compared to results previously achieved by manual operators.
8. Human resource management: Text mining techniques are also used to manage human resources strategically, mainly with applications aiming at analyzing staff’s opinions, monitoring the level of employee satisfaction, as well as reading and storing CVs for the selection of new personnel. In the context of human resources management, the TM techniques are often utilized to monitor the state of health of a company by means of the systematic analysis of informal documents.
9. Customer Relationship Management (CRM): Text mining in CRM domain is most widely used in the areas related to the management and analysis of the contents of client’s messages. This kind of analysis often aims at automatically rerouting specific requests to the appropriate service or at supplying immediate answers to the most frequently asked questions. Services research has emerged as a green field area for application of advances in computer science and IT. CRM practices, particularly contact centers (call centers) in our context, have emerged as hotbeds for application of innovations in the areas of knowledge management, analytics, and data mining. Unstructured text documents produced from a variety of sources in today contact centers have exploded in terms of the sheer volume generated. Companies are increasingly looking to understand and analyze this content to derive operational and business insights. The customer, the end consumer of products and services, is receiving increased attention. Business analytics applications revolving around customers have led to emergence of areas like customer experience management, customer relationship management, and customer service quality. These are becoming critical to competitive growth, and sometimes even, survival. Applications with such customer focus are most evident in services companies especially CRM practices and contact centers.
10. Market Analysis: Text mining in Market Analysis is used mainly to monitor customer’s opinion to identify new potential customers, analyze competitors and determine the organization’s image by analyzing press reviews and other relevant sources. Most of the organization indulge in tele-marketing and e-mail activities to acquire new customers. With the introduction of text mining systems organizations are able to answer the queries related to more complex market scenarios. Data mining technology have helped us in extracting useful information from various databases. Data warehouses turned out to be successful for numerical information, but failed when it came to textual information. The 21st century has taken us beyond the limited amount of information on the web. This is good in one way that more information would provide greater awareness, and better knowledge. The knowledge of marketing information is available on the web by means of industry white papers, academic publications relating to markets, trade journals, market news articles, reviews, and even public opinions when it comes down to customer requirements.

Target marketing: Find clusters of “model” customers who share the same characteristics: interest, income level, spending habits, etc.

Determine customer purchasing patterns over time: Conversion of single to a joint bank account: marriage, etc.

Cross-market analysis: Associations/co-relations between product sales, Prediction based on the association information, Finance planning and asset evaluation

1. Warranty or insurance claims, diagnostic medical interviews, etc.: In certain business areas, the bulk of the information available is in an undefined textual form. For example, during warranty or insurance claims, claimant will be interviewed by an insurance agent and he will take note of all the details related to the claim in the form of a brief description. Similarly during patient medical interviews, the attendant will take down a brief description of the patient’s health issues or when you take your vehicle for repairs to the service station, the attendant will take down some notes about the issues you highlight and what needs to be repaired. These notes are then collected electronically and are input into the text mining systems. This information can be exploited to identify common cluster of problems and complaints on certain vehicles, etc. Similarly in the medical field useful information can be extracted from the collected open-ended descriptions about patient’s disease symptoms, which could be helpful in actual medical diagnosis.
2. Search and information retrieval (IR): Storage and retrieval of text documents, including search
3. engines and keyword search.
4. Document clustering: Grouping and categorizing terms, snippets, paragraphs, or documents, using
5. data mining clustering methods.
6. Document classification: Grouping and categorizing snippets, paragraphs, or documents, using
7. data mining classification methods, based on models trained on labeled examples.
8. Web mining: Data and text mining on the Internet, with a specific focus on the scale and
9. interconnectedness of the web.
10. Information extraction (IE): Identification and extraction of relevant facts and relationships from
11. unstructured text; the process of making structured data from unstructured and semistructured text.
12. Natural language processing (NLP): Low-level language processing and understanding tasks (e.g.,
13. tagging part of speech); often used synonymously with computational linguistics.
14. Concept extraction: Grouping of words and phrases into semantically similar groups.