Implementing Text Analytics with Enterprise Resource Planning

Implementing Text Analytics with Enterprise Resource Planning

Surjit Singh Bawa

Solutions Architect
Carmel, Indiana, USA
Surjit.sb@gmail.com

Abstract - **The increasing volume of unstructured data in organizations has led to the adoption of text analytics to extract valuable insights. In this article, we discuss the implementation of text analytics with Enterprise Resource Planning (ERP) systems to improve decision-making processes. We provide recommendations for overcoming these challenges and offer a roadmap for organizations to follow when implementing text analytics in their ERP systems.**

Keywords - Intelligent Enterprise, ERP, Text Analytics, Machine Learning.

I. INTRODUCTION TO TEXT ANALYTICS IN ENTERPRISE RESOURCE PLANNING (ERP)

Enterprise Resource Planning (ERP)[1] is used by organizations to integrate their business processes, manage resources, and streamline operations. It enables organizations to manage and optimize their business functions such as finance, inventory, human resources, customer relationship management, and more.

With the increasing volume of unstructured data such as documents, customer feedback, social media comments, email, and other text data generated by businesses daily, there is a growing need for tools to help extract insights from this data. This is where Text Analytics comes in.

Text Analytics [2] is a technology that uses natural language processing (NLP) and machine learning algorithms to analyze and extract valuable information from unstructured text data. It can help organizations understand customer needs and preferences, identify emerging trends, improve product and service offerings, and more.
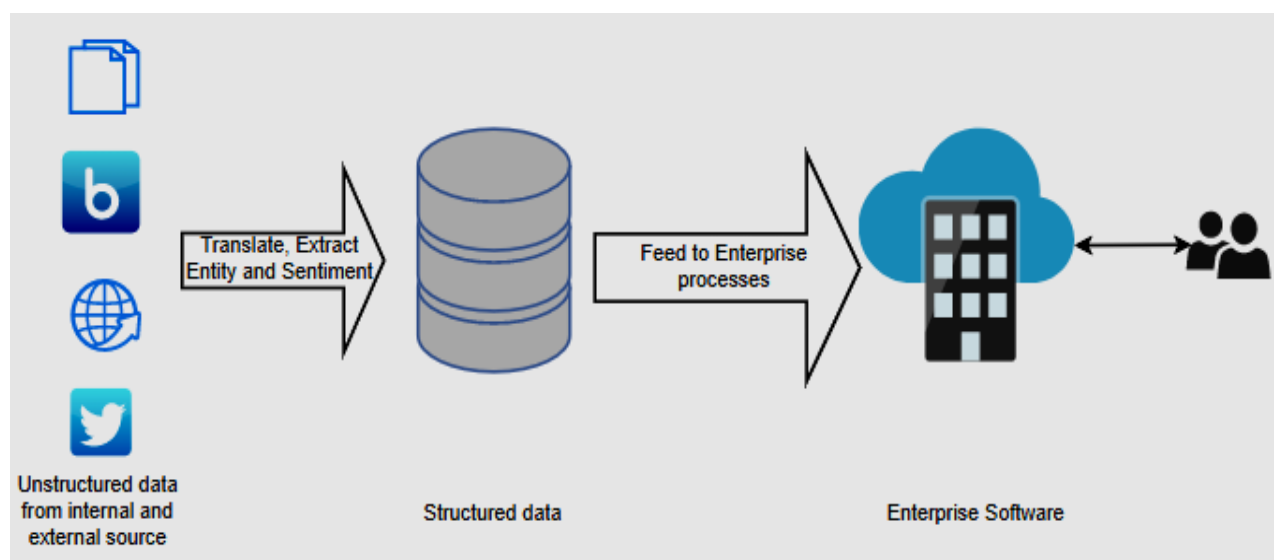

Figure 1. Retrieving unstructured data

Integrating Text Analytics with ERP systems as shown in Figure 1 can provide organizations with a comprehensive view of their business operations, including customer feedback, supplier performance, and employee sentiment. It can help organizations make informed decisions, enhance customer satisfaction, and drive business growth.

In this paper, we will explore the role of Text Analytics in ERP, its key challenges, strategies for successful implementation, benefits of using Text Analytics in ERP, real-life use cases, and the future of Text Analytics in ERP.

II. THE ROLE OF TEXT ANALYTICS IN ERP

Text Analytics with Enterprise Resource Planning (ERP) as shown in Figure 2 enables organizations to leverage unstructured text data to gain insights into their business operations, customers, and competitors. Text Analytics can help organizations in the following ways:

Customer Insight [3]: Text Analytics can help organizations understand customer needs and preferences by analyzing customer feedback, social media comments, and other unstructured text data. This insight can be used to improve customer service, develop targeted marketing campaigns, and optimize product and service offerings.
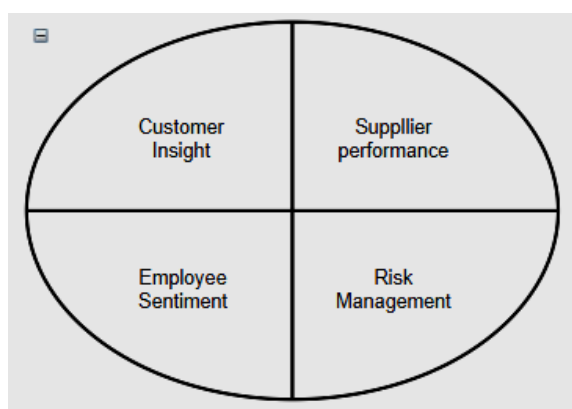

Figure 2. Text analytics and ERP Process

Supplier Performance [4]: Text Analytics can help organizations analyze supplier performance by extracting information from unstructured text data such as supplier invoices, purchase orders, and contracts. This can help organizations identify areas for improvement, negotiate better prices, and reduce supply chain risks.

Employee Sentiment [5]: Text Analytics can help organizations analyze employee sentiment by extracting information from unstructured text data such as employee surveys, social media comments, and performance reviews. This insight can be used to improve employee engagement, retention, and productivity.

Risk Management [6]: Text Analytics can help organizations identify potential risks by analyzing unstructured text data such as news articles, social media comments, and regulatory filings. This insight can be used to mitigate risks, comply with regulations, and maintain reputation.

Competitive Intelligence: Text Analytics can help organizations gain a competitive advantage by analyzing unstructured text data such as industry reports, news articles, and social media comments. This insight can be used to identify emerging trends, monitor competitor activity, and develop strategic plans.

III. KEY CHALLENGES OF IMPLEMENTING TEXT ANALYTICS IN ERP

Although Text Analytics can provide valuable insights, implementing it with Enterprise Resource Planning (ERP) systems can pose significant challenges [7]. Here are some of the key challenges:

- *Data Quality:* Text Analytics relies on high-quality data to provide accurate insights. However, unstructured text data can be challenging to clean, process, and normalize. Poor data quality can result in inaccurate insights, leading to incorrect decisions.

- *Data Integration:* Integrating unstructured text data with ERP systems can be complex, especially if the data is stored in disparate systems. Ensuring data consistency, accuracy, and completeness can be challenging, leading to integration issues.

- *Data Volume:* Text data can be voluminous, making it challenging to store, process, and analyze. Organizations need to invest in robust infrastructure and scalable solutions to handle large volumes of data.

- *Data Security:* Text data may contain sensitive information such as customer details, financial information, and proprietary business information. Organizations must ensure that data security measures are in place to protect data from unauthorized access, breaches, and cyber threats.

- *Change Management:* Implementing Text Analytics in ERP systems requires changes in business processes, workflows, and user roles. Organizations must invest in change management initiatives to ensure that stakeholders are prepared and willing to adopt these changes.

Addressing these challenges requires a comprehensive approach that includes robust data management processes, technology investments, organizational change management, and talent acquisition strategies. By addressing these challenges, organizations can harness the power of Text Analytics to gain valuable insights and drive business growth.

IV. UNSTRUCTURED DATA SOURCES

There are multiple sources of unstructured data as shown in Figure 3, these are segregated in two groups – Internal and External. Internal data is generated by the Organization and External data is generated by Customers, Employees and Supplier on platforms that are outside the organization.

Figure 3. Source of Data

Documents: Every enterprise will generate documents related to Contracts, Sales, Invoices, Product design, Product Quality, Manufacturing, and, Purchase. These documents contain important information related to Products that the company Sells, Manufactures and Procures.

E-Mails: Most of the communication in today's digital world is done through e-mails between Customer and Sales teams, Supplier and Purchase teams, HR and Employees. These emails can contain information on Product feedback, Product quality issues, Supplier issues, employee feedbacks.

Customer Support: Support teams are important interface between company and Customer, they capture information on Product issues, returns, Purchase issues and feedback. They are generally logged in support systems manually by Agents.

Surveys: Many companies undertake surveys, to identify issues faced by Employees and Customers, these surveys contain information that is added manually by the users, which can be extracted and consumed by Text analytics.

Customer Feedback: Customers will generally leave feedback on the website from where they purchased the product, this could be E-commerce, third-party sales, or Direct to customer sales, All this information is valuable that is related to the product.

Reviews: Employees generally write feedback or share their experience on Social media or other platforms about the company, this data is generally entered manually by current or ex-employees.

Social Media: Social media plays as an important role when marketing products, Customers will share their experience and might also suggest products to others on social media, this data is can be harnessed and used for Marketing, connecting with customers and identifying potential issues with product.

V. TEXT ANALYTICS METHODOLOGIES

After identifying source of unstructured data, it is important to understand the steps to Analyze and consume this data.

Some of the most important steps as shown in figure 4 are explained below.

Bag-of-Words: In text analytics, a bag-of-words model is a way of representing text data as a collection of individual words, or "tokens," without considering their order or context. It is called a "bag" of words because the model only keeps track of the frequency of each word in the document, not their position or relationship to other words. To create a bag-of-words representation, the text is first preprocessed to remove punctuation, numbers, and stop words (common words such as "the," "and," "a," etc. that do not carry much meaning).

Then, each word in the text is counted and represented as a separate feature in a data matrix, with the frequency of each word being its value in the corresponding column. This matrix can then be used as input to machine learning algorithms for tasks such as sentiment analysis, text classification, or topic modeling. While bag-of-words models are simple and computationally efficient, they have several limitations. They do not capture the meaning of phrases or sentences, they treat all words as equally important, and they are not good at handling misspellings or

rare words. More advanced NLP techniques, such as word embedding or contextualized representations, are often used

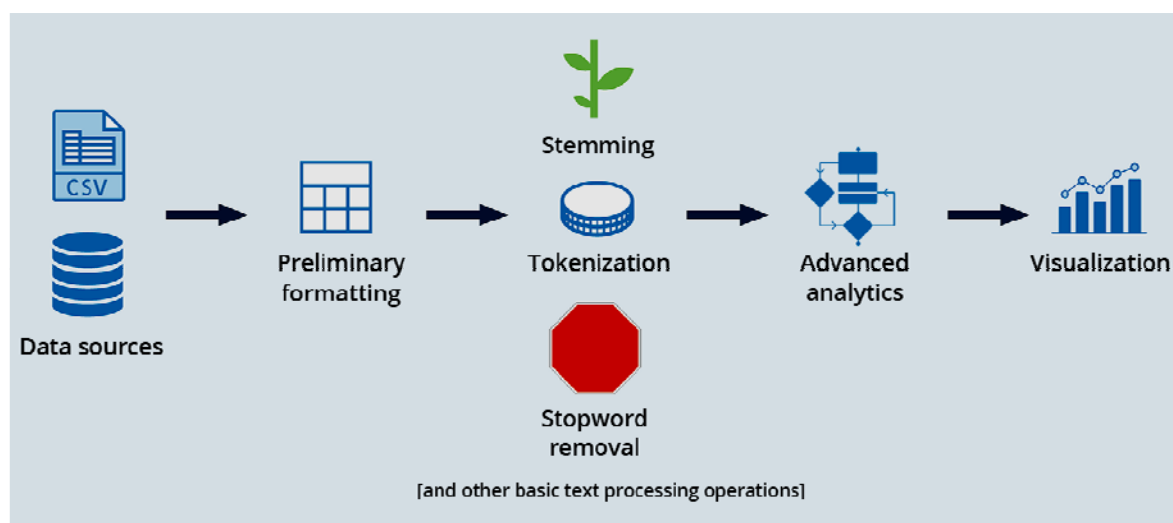to address these limitations and improve the accuracy of text analysis tasks.

Figure 4. Text analytics Methodologies[10]

Stemming: Stemming is a text analysis technique that involves reducing words to their root or base form in order to group together words that have similar meanings. It is commonly used in natural language processing (NLP) to preprocess text data for analysis. The process of stemming involves removing suffixes and prefixes from words so that they can be matched to other words with the same root. For example, the words "running," "runners," and "run" can be stemmed to "run." Stemming can be useful for a variety of NLP tasks, including information retrieval, text classification, and sentiment analysis. By reducing the number of unique words in a text corpus, stemming can also help to improve the efficiency of machine learning algorithms that are used to analyze large volumes of text data. However, it's important to note that stemming can also introduce errors into text analysis, as some words may be stemmed incorrectly and lose their original meaning. Therefore, it's important to use stemming carefully and in conjunction with other text analysis techniques to ensure the accuracy of the results.

Tokenization: In text analytics, tokenization is the process of breaking down a piece of text into individual units called tokens. These tokens can be words, phrases, or even characters, depending on the level of granularity required for the analysis. Tokenization is an important step in many natural language processing tasks, such as text classification, sentiment analysis, and machine translation. It helps to standardize the input data and make it easier to process and analyze. The tokenization process typically involves the following steps:

- Removing any unwanted characters, such as punctuation marks or numbers.
- Splitting the text into individual words or phrases.
- Converting all the tokens to lowercase or uppercase to standardize the text.
- Removing any stop words, which are common words that do not carry much meaning, such as "the," "and," or "a."

There are several tools and libraries available for tokenization in different programming languages, such as NLTK (Natural Language Toolkit) in Python or CoreNLP in Java.

Tokenization is an important preprocessing step in text analytics because it enables further analysis of the text, such as counting the frequency of each token, identifying the most common words, or generating features for machine learning models.

Stop-Word Removal: In text analytics, stop-word removal is a common technique used in preprocessing text data. Stop-words are words that occur frequently in a language but do not carry much meaning, such as "the," "and," "a," etc. They are often removed from text data because they can add noise to the analysis and do not provide much useful information. The process of stop-word removal involves identifying a list of stop-words in a given language and removing them from the text data. This can be done manually, by creating a custom list of stop-words specific to the domain or application, or automatically, by using predefined lists of stop-words available in various NLP libraries. Stop-word removal can improve the accuracy

5.4

and efficiency of various text analysis tasks, such as text classification, sentiment analysis, and topic modeling. By removing commonly occurring words, stop-word removal can help to focus the analysis on the more meaningful and relevant words in the text.

However, it is important to note that stop-word removal is not always necessary or appropriate for every text analysis task. In some cases, stop-words may carry important information or context, and removing them can lead to a loss of valuable information. Therefore, it's important to evaluate the impact of stop-word removal on the specific task at hand and make an informed decision about whether to include or exclude stop-words from the analysis.

Cluster Analysis: Cluster analysis is a text analytics technique that involves grouping similar documents or words together into clusters based on their content or similarity. The goal of cluster analysis is to identify patterns or structures within large sets of text data, which can be useful for various text analysis tasks such as text classification, topic modeling, or information retrieval. There are several approaches to cluster analysis in text analytics, such as hierarchical clustering, K-means clustering, or density-based clustering. Each approach has its own strengths and weaknesses, depending on the specific task and dataset.

In *Hierarchical Clustering*, the text data is recursively divided into smaller clusters based on their similarity, until each cluster contains only a few documents or words. This approach can be useful for exploring the structure of the data and identifying meaningful subgroups, but it can also be computationally expensive and may not scale well to very large datasets.

In *K-Means Clustering,* the text data is divided into a predefined number of clusters based on their similarity, with each cluster represented by a centroid (or center) of the cluster. This approach can be more efficient than hierarchical clustering and can work well for larger datasets, but it requires the number of clusters to be specified in advance, which may not always be known in advance.

In *Density-Based Clustering*, the text data is clustered based on their density and connectivity, with clusters defined as areas of high density separated by areas of lower density. This approach can be useful for identifying clusters of different sizes and shapes, but it can also be sensitive to the choice of parameters and may not work well for very sparse or high-dimensional data.

Overall, cluster analysis is a powerful tool in text analytics for discovering patterns and structures in large sets of text data, and it can be used in combination with other techniques such as dimensionality reduction, feature

extraction, or sentiment analysis to gain deeper insights into the data.

As an example text analytics can be used to generate Word intensity and relations as shown in figure 5 and 6.
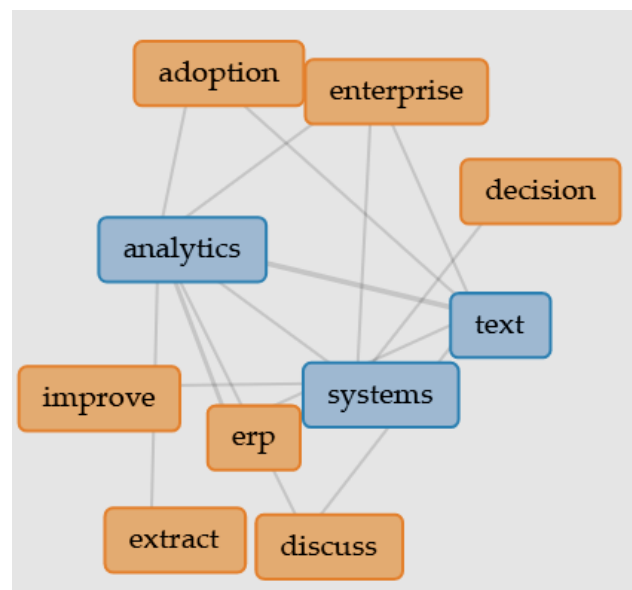

Figure 5. Word intensity

Figure 6. Word relations

VI. STRATEGIES FOR SUCCESSFUL IMPLEMENTATION OF TEXT ANALYTICS IN ERP

Implementing Text Analytics [8] with Enterprise Resource Planning (ERP) systems can be challenging, but with the right strategies, organizations can successfully implement this technology. Here are some strategies for the successful implementation of Text Analytics in ERP:

- *Define Clear Objectives*: Organizations should clearly define their objectives for implementing Text Analytics in

ERP systems. This includes identifying the data sources, use cases, and expected outcomes.

- *Develop a Data Strategy:* Organizations should develop a comprehensive data strategy that includes data quality, integration, storage, and security. This strategy should align with the organization's overall data management strategy.

Invest in the Right Technology: Organizations should invest in technology that can handle large volumes of text data, provide accurate insights, and integrate with ERP systems. This includes Natural Language Processing (NLP) tools, machine learning algorithms, and visualization tools.

- *Start Small and Scale:* Organizations should start with a small pilot project to test the technology and refine the implementation strategy. Once the pilot is successful, they should scale up gradually, taking into account the organizational impact, change management, and resource requirements.

- *Collaborate Across Functions:* Implementing Text Analytics in ERP systems requires collaboration across functions, including IT, data management, business analytics, and business operations. Organizations should foster cross-functional collaboration to ensure that the implementation aligns with business goals and objectives.

- Measure and Monitor Progress: Organizations should establish performance metrics to measure the success of Text Analytics implementation. They should also monitor progress regularly, identify gaps and areas for improvement, and take corrective action as necessary.

VII. REAL-LIFE USE CASES OF TEXT ANALYTICS IN ERP

Text Analytics can be applied in a variety of ways within Enterprise Resource Planning (ERP) systems to provide valuable insights and improve business processes [9]. Here are some real-life use cases of Text Analytics in ERP:

- *Sentiment Analysis for Customer Feedback:* Text Analytics can be used to analyze customer feedback and sentiment to gain insights into customer needs and preferences. For example, a manufacturer can use Text Analytics to analyze customer reviews of their products to identify areas for improvement.

- *Fraud Detection and Prevention*: Text Analytics can be used to analyze unstructured text data, such as emails and chat logs, to identify potential fraud and other financial crimes. For example, an insurance company can use Text Analytics to identify fraudulent claims by analyzing the text data associated with those claims.

- *Predictive Maintenance:* Text Analytics can be used to analyze text data from maintenance logs and other sources to predict equipment failures and maintenance needs. For example, a power plant can use Text Analytics to predict when a piece of equipment is likely to fail, enabling them to perform maintenance proactively.

- *Quality Control and Inspection:* Text Analytics can be used to analyze text data from inspection reports to identify trends and patterns related to quality issues. For example, a manufacturer can use Text Analytics to identify the root cause of a quality issue by analyzing the text data associated with the production process.

- *Supply Chain Optimization:* Text Analytics can be used to analyze unstructured text data from suppliers, logistics providers, and other sources to optimize supply chain operations. For example, a retailer can use Text Analytics to analyze supplier invoices to identify discrepancies and negotiate better terms.

- *Employee Engagement and Satisfaction:* Text Analytics can be used to analyze employee sentiment and engagement levels to identify areas for improvement. For example, a human resources department can use Text Analytics to analyze employee feedback from surveys and other sources to identify trends and patterns related to employee satisfaction.

VIII. FUTURE OF TEXT ANALYTICS IN ERP

The future of Text Analytics in Enterprise Resource Planning (ERP) is promising. Here are some potential advancements that can be expected in the future:

- *Increased Integration with AI and Machine Learning:* Text Analytics can be integrated with artificial intelligence (AI) and machine learning (ML) technologies to enhance its capabilities. This can enable ERP systems to automatically identify patterns and trends in unstructured text data and make predictions based on that data.

- *Enhanced Natural Language Processing (NLP)*: As NLP technology continues to advance, Text Analytics will become more sophisticated in its ability to analyze and interpret natural language. This can enable ERP systems to understand more complex language structures, identify sentiment more accurately, and recognize more subtle nuances in text data.

- *Greater Emphasis on Contextual Analysis:* As Text Analytics technology continues to evolve, there will be a greater emphasis on contextual analysis. This can enable ERP systems to better understand the context in which text data is being used, and make more accurate and relevant predictions based on that data.

- Improved Data Visualization: As organizations continue to collect more and more data, it becomes increasingly important to be able to visualize that data in a meaningful way. Text Analytics can be integrated with data visualization tools to provide interactive and visually appealing representations of text data, enabling organizations to gain insights at a glance.

- Greater Focus on Privacy and Security: As the amount of data collected by ERP systems continues to grow, there will be a greater focus on privacy and security. Text Analytics technology will need to be developed with privacy and security considerations in mind, to ensure that sensitive data is protected and used ethically.

Overall, the future of Text Analytics in ERP is likely to be characterized by increased sophistication, integration with other technologies, and a greater emphasis on privacy and security. By leveraging Text Analytics, organizations can gain valuable insights from unstructured text data, enabling them to make better decisions, reduce costs, and drive business growth.

IX. CONCLUSION AND RECOMMENDATIONS FOR ADOPTING TEXT ANALYTICS IN ERP.

In conclusion, Text Analytics can provide a significant competitive advantage to organizations by providing insights and improving business processes within Enterprise Resource Planning (ERP) systems. However, there are also key challenges that must be overcome in order to successfully implement Text Analytics in ERP.

Organizations considering adopting Text Analytics in ERP should take the following recommendations into consideration:

- Identify Clear Business Objectives: Before implementing Text Analytics in ERP, it is important to identify clear business objectives that align with the overall goals of the organization. This can help ensure that Text Analytics is being used in a way that adds value and drives business growth.

- Invest in the Right Tools and Technologies: Implementing Text Analytics in ERP requires the right tools and technologies, including software and hardware. It is important to invest in the right tools that are aligned with the specific needs and goals of the organization.

- Ensure Data Quality: Text Analytics relies on high-quality data. It is important to ensure that data is clean, accurate, and complete before implementing Text Analytics in ERP.

- Develop a Strong Data Governance Framework: Implementing Text Analytics in ERP requires a strong data governance framework that includes policies and procedures for data access, security, and privacy. This can help ensure that data is being used ethically and responsibly.

- Invest in Training and Education: Text Analytics is a complex technology that requires specialized skills and knowledge. It is important to invest in training and education to ensure that employees have the skills and knowledge necessary to use Text Analytics effectively.

By following these recommendations, organizations can successfully implement Text Analytics in ERP and gain a competitive advantage by leveraging valuable insights from unstructured text data.

REFERENCES

[1] Klaus, Helmut, Michael Rosemann, and Guy G. Gable. "What is ERP?." Information systems frontiers 2 (2000): 141-162.

[2] Rockwell, Geoffrey. "What is text analysis, really?." Literary and linguistic computing 18, no. 2 (2003): 209-219.

[3] Gallagher, Conor, Eoghan Furey, and Kevin Curran. "The application of sentiment analysis and text analytics to customer experience reviews to understand what customers are really saying." International Journal of Data Warehousing and Mining (IJDWM) 15, no. 4 (2019): 21-47.

[4] Sabbagh, Ramin. "Semantic Text Analytics Technique for Classification of Manufacturing Suppliers." (2018).

[5] Goldberg, David, and Nohel Zaman. "Text analytics for employee dissatisfaction in human resources management." (2018).

[6] Haile, Issayas M. "Data Analytics in Financial Institutions: How Text Analytics Can Help in Risk Management." PhD diss., Colorado Technical University, 2020.

[7] Ittoo, Ashwin, and Antal van den Bosch. "Text analytics in industry: Challenges, desiderata and trends." Computers in Industry 78 (2016): 96-107.

[8] Anandarajan, Murugan, Chelsey Hill, and Thomas Nolan. "Practical text analytics." Maximizing the Value of Text Data.(Advances in Analytics and Data Science. Vol. 2.), Springer (2019): 45-59.

[9] Miner, Gary, John Elder IV, Andrew Fast, Thomas Hill, Robert Nisbet, and Dursun Delen. Practical text mining and statistical analysis for non-structured text data applications. Academic Press, 2012.

[10] https://www.softwareadvice.com/resources/what-is-text-analytics/.