Optimizing Engineering Change Management Through ECR Data Analysis and Mining Techniques

Zheng Xinyu and He Xin
Business School, Renmin University of China, Haidian District, Beijing, P.R.China.
[1]xinyu@rmbs.ruc.edu.cn, [2]hexin@rmbs.ruc.edu.cn

Correspondence should be addressed to Zheng Xinyu : xinyu@rmbs.ruc.edu.cn

Article Info
Journal of Enterprise and Business Intelligence (https://anapub.co.ke/journals/jebi/jebi.html)
Doi: https://doi.org/10.53759/5181/JEBI2025022.
Received 16 March 2025; Revised from 18 May 2025; Accepted 30 June 2025.
Available online 05 October 2025.

Abstract – In the current field of industrial engineering, Engineering Change Management (ECM) has become a vital factor for preserving product quality and encouraging new ideas. This paper aims to understand how the management of the ECM processes is done at a global automobile firm particularly in the case of Engineering Change Requests (ECR). This study reveals that over 70 new ECRs are created daily, and in the past 10 years, over 120,000 ECRs have been recorded; it stresses the need for better Information Systems (IS) to tackle this problem. This research will employ the Knowledge Discovery in Databases (KDD) framework and text mining to analyze historical ECR data in a bid to make recommendations that should assist to enhance ECM effectiveness and decision making. The findings showed that text mining process consists of data extraction and data pre-processing, document segmentation and tokenization, removal of stopwords, stemming of words and word vector generation yielding 5,783 unique terms or keywords. Through cluster analysis, it was possible to identify common problem areas with some damage in the ECRs. We included keywords frequencies and clusters for Projects A, B and C; which aided in identifying specific project issues and trends with each project. For instance, in Project A, the term frequency and potential issues were pointed out; however, Projects B and C had differences in the term frequency distribution, which helped to pinpoint the common problems and direct further ECM enhancements.

Keywords – Engineering Change Management (ECM), Knowledge Discovery in Databases (KDD), Engineering Change Requests (ECR).

I. INTRODUCTION

Knowledge Discovery in Databases (KDD) [1], which came into existence in the 1980s, is an effective approach to understanding big data, machine learning, and artificial intelligence. The most conspicuous definition of KDD was posted in 1996 by Fayyad, Piatetsky-Shapiro and Smyth [2] as the identification of hidden patterns in the data that are significant, novel, perhaps useful, and comprehensible. This definition can also be extended to the definition of the concept termed as "Data Mining" (DM) or text mining. In fact, in the present literature concerning Data Mining (DM) [3] and KDD, the two terms are interchangeably or with not much distinction. Classical KDD techniques [4] define DM as a step in the KDD procedure that involves extracting knowledge. This step also includes selecting and preparing suitable data from different sources, as well as accurately interpreting the outcomes of the mining process. Common DM techniques encompass text mining, concept description, time-series analysis, association rule mining, clustering analysis, classification and prediction [5]. Over the past 20 years, the area of KDD has gained significant attention across various fields, including telecommunications, banking, marketing, and scientific analysis.

In the present-day business, there are several Engineering Changes (ECs) [6] that involve altering the fits, functionalities, materials, dimensions, and other aspects of a product and its elements after the plan has been approved. In 2005, Ford, GM, and DaimlerChrysler solely focused on quantifying the internal ECs within their supply chain [7]. The combined annual revenue for the three enterprises was 350,000 ECs. The projected cost for each EC, including hidden expenditures, was $50,000 [8]. The significant magnitudes and expenses of ECs underscore the need of engineering change management (ECM). Over the past few decades, numerous companies have prioritized enhancing their efficiency in implementing electronic commerce (ECs). In order to streamline the bureaucratic EC process, Workflow Management Systems (WfMS) [9] were implemented to facilitate the smooth passage of ECs. Consequently, a substantial amount of information regarding the process is kept [10]. Nevertheless, it has become evident that efficiency alone is not the sole critical factor for success. Companies must also enhance their effectiveness, enabling them to implement appropriate modifications at the optimal stage of the design process [11]. However, corporations receive limited assistance in their decision-making process, and the individuals involved in the process typically perform these activities manually [12]. Locating information in modern systems can be exceedingly challenging, particularly when the data is dispersed across various databases and necessitates the use of numerous applications or systems. Therefore, a significant amount of time during the EC process is dedicated to information retrieval [13].

In organizations that engage in the design and production of intricate items, particularly those that are customized, it is common for alterations and adjustments to occur during the product's development. Several of these modifications are explicitly instigated by the client as fresh demands, or by the organization as altered specs or production adjustments. Engineering Change Requests (ECRs) can be made prior to, during, or after production. The later these occurrences happen, the greater the amount of time and effort needed to carry them out [14]. Modifications may also arise throughout the utilization of a product as a result of design flaws, the need for replacement parts, or enhancements made towards the end of the product's lifespan. The usage of ECRs affects many departments in a firm such as engineering, sales, production, purchasing, inventory control, accountancy, and many others. Since speed to market is one of the key factors of competition, ECRs are better to be managed efficiently and accurately. However, the procedure can lead to important costs due to waste, delay, new acquisitions and replacement of the components stock [15].

To manage the requirements in Product Development Processes (PDPs), Engineering Change Requests (ECR) is one of the solutions. The concept of ECM encompasses the elements of integration of development activities, inspection and other related activities [16]. The scope of this task includes overseeing and supervising the procedures involved in implementing and tracking changes inside an organization [17]. Engineering changes are regarded as alterations to previously distributed papers or components [18]. The products and their design and production methods eventually reflect these factors [19]. Iterative methods are particularly required in complex and large-scale undertakings. However, the mere act of generating such a request requires time and other resources and does not contribute value. We are of the opinion that ECM, particularly the establishment of ECRs, can be enhanced by bringing to light concealed, yet valuable, knowledge. That information can be located in the ECM history. Enhancements in knowledge generation and data administration can expedite the achievement of development outcomes. Organizations can gain a competitive edge by recognizing that the outcome of a PDP [20] is not just the physical product, but also the data and knowledge about how to generate future products. These artifacts, which are created during the development process, are commonly gathered in a unified product information model. Thus, it is clear that there are apparent advantages to be gained from implementing KDD. KDD can be utilized to identify parallels between ECR and extract patterns.

This study seeks to apply the KDD framework and Text Mining techniques to convert ECR data into valuable information that would improve the efficacy and decision-making of ECM processes. This research aims to fill the existing research gap in the current management practices and aims to establish the possibility of enhancing the effectiveness of change management programs in global automobile firms through the use of advanced data study tools. The rest of the paper has been arranged as follows: Section II reviews related works on Engineering Change Management (ECM), Knowledge Discovery in Databases (KDD), and Engineering Change Requests (ECR). Section III provides a review of the data and methods concerning the KDD framework, ECM, and data collection. Section IV and V provide a critical survey and discussion of the results obtained in this research. Lastly, Section VI provides a conclusion to the research, and recommends future research works on data science solutions in improving ECM practices to be more responsive, effective, and effective for various organizational goals.

II. RELATED WORK

Knowledge Discovery in Databases (KDD) [21]-[25] is the systematic process of deriving valuable and beneficial knowledge from a dataset. The process of knowledge extraction is crucial for obtaining vital information, and indiscriminate data mining can potentially result in meaningless patterns, posing significant risks. **Fig 1** illustrates the sequential process of knowledge discovery. The KDD process can be structured using the process model developed by Kurgan and Musilek [26], Rotondo and Quilligan [27], and Wu et al. [28]. The process is fundamentally iterative, beginning with the selection of data, followed by preprocessing to change it into a more digestible format. DM algorithms are then applied to the transformed data, and the results are interpreted. Incorporating Business Understanding as the initial step, Plotnikova, Dumas, and Milani [29] and Wiemer, Drowatzky, and Ihlenfeldt [30] extended the process, aligning it with the CRISP-DM model's phase [31]. It is crucial for the successful execution of a KDD project and acts as a guide for applying data mining projects. The model encompasses all essential phases and tasks, together with their interconnections, presented as a cyclic process model.

In [32], engineering changes (EC) have been considered to be an indispensable component of production design. They are utilized to enhance and modify items or to achieve a predetermined state of the product that has not been attained due to a specific issue. However, ECs also occupy a significant portion of the development capacity in industry, ranging from 30% to 50% and occasionally even up to 70% [33]-[35]. This highlights the imperative of effectively managing ECs in order to avoid either stifling its growth through excessive adjustments or squandering the opportunity for a successful product. In a case study conducted by Marks et al. [36] and Eisenhardt and Martin [37], five solutions for dealing with ECs were identified: prevention, frontloading, effectiveness, efficiency, and learning. Nevertheless, EC Managers have a deficiency in understanding the mechanisms or patterns within the organization that are necessary to properly implement the strategy. To enhance the efficiency of ECs, it is necessary to understand the company-specific mechanisms of previous ECs that had a lengthy lead time. Examples of such mechanisms or patterns could include ECs that involve more than five departments and occur quickly after a Design Freeze, often with a significant lead time. Typically, these patterns are not present in firms, which means that the measures chosen typically do not result in a desire to enhance the ECM.

Text mining [38], also known as Knowledge Discovery from Text (KDT) [39], was first introduced in [40]. It involves using machines to analyze text. This approach incorporates methodologies from information extraction, natural language processing (NLP), information retrieval, and integrates them with the systems and techniques of statistics, KDD, machine learning, and DM.

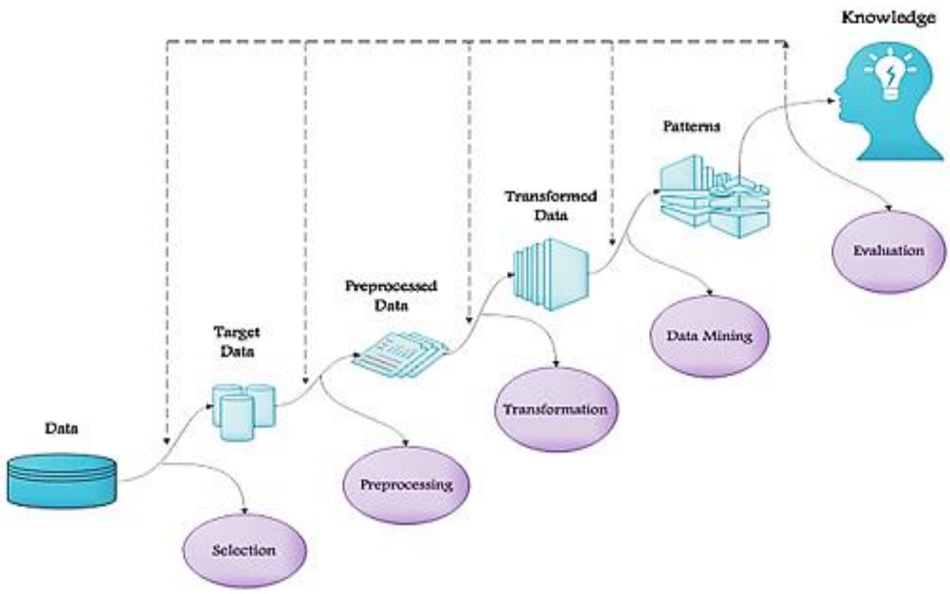

Fig 1. Knowledge Discovery Steps.

Therefore, one follows a comparable approach to the KDD process, but instead of analyzing data in general, the focus is specifically on text texts. These findings give rise to additional inquiries regarding the employed data mining techniques [41]. An issue arises when we are faced with unstructured data sets from the perspective of data modeling. Zheng et al. [42], Haddow et al. [43], Dunne et al. [44], Bizer et al. [45] focuses on addressing issues related to text representation, categorization, clustering, information extraction, and the identification and modeling of concealed patterns. According to Abu-Salih [46], the selection of characteristics and the impact of domain knowledge and domain-specific techniques are crucial in this context. Consequently, it is typically required to modify existing DM algorithms so as to analyze text data. To accomplish this, individuals often depend on the expertise and findings of studies in data retrieval, NLP, and information extraction.

III. DATA AND METHOD

KDD Framework

KDD can be considered as the general framework for learning from large databases with the use of AI principles, database management, statistical, and machine learning approaches. Originally introduced in the late 1980s, KDD has been expanded through several stages which are critical in helping to convert large amounts of raw data into useful knowledge [47]. As stated by Rudolph [48], KDD is described as the nontrivial procedure of finding understandable, valid, possibly useful, and novel patterns within data sets. In [49], it is clear that the KDD process is iterative and includes selecting data, formulating the data, altering the data, DM, and interpreting the data. The first step in the KDD procedure is the selection of the data source, which involves defining the necessary datasets and obtaining them from different sources in the organization.

Table 1. Attributes of ECR Data

Attribute	Description	References
Product Design	Modifications in product design specifications	[50, 51, 52, 53, 54]
Materials	Changes in materials used	[55, 56, 57]
Dimensions	Alterations in product dimensions	[58]
Functionality	Adjustments in product functionality	[59, 60, 61]
Date of Creation	Timestamp when the ECR was created	[62]
Responsible Department	Department or team responsible for the ECR	[63, 64]

We investigate a large German automobile company with a multifaceted, federal, and decentralized ECM process that is backed by information systems. This manufacturer produces several automobile products every year, which requires a strong product development department. Due to the large scale of operations, approximately 70 new ECRs are created daily, which contain essential information outlined in **Table 1** that includes features like design change, material, dimension, and functional change. Data preprocessing in the setting of the KDD process entails data transformation, data integration, and data cleaning to advance the quality and compatibility of the data. Outlier detection and normalization are used to make the dataset ready for the further data mining process. Classification and clustering systems are then used to analyze the large data set of over 120000 ECRs collected over the last decade. These findings are essential in analyzing changes in the creation and resolution of ECRs and in the use of past data to shape future change management.

ECM and Its Challenges

ECM is a key process in industries that are elaborated in the advancement of complex products that assist in the management of Engineering Change Requests (ECRs). ECRs include changes in product design, material, size, and use, which are carried out internally and by outside parties. ECRs are crucial for maintaining product integrity, shortening development cycles and controlling the costs of change [65]. In practice, therefore, the concept of ECM within our study setting refers to a process that is defined in advance and applied through the use of an IT scheme to address the high ECR throughput. The structure also helps to guarantee that changes are assessed thoroughly, authorized according to a set of guidelines, and integrated into the different product lines without incident. Interviewing 16 out of 40 guideline-based contributors, challenges in ECM have emerged, highlighting the difficulty of incorporating newly generated ECRs with historical data for a more comprehensive improvement process.

Data Collection and Study Context

In this research, we concentrate on a global automobile firm with a large research and development division and a diverse range of products. Thus, the corporation handles about 70 new ECRs per day regarding various products, which indicates the nature and scope of engineering within the company. Semi-structured interviews were conducted following guidelines with 16 participants focusing on the qualitative aspects of ECM practices, barriers, and potential enhancements. These interviews were designed to focus on current practice, to gain an understanding of patterns in ECR creation and resolution, and to determine the use of previous ECR information [66]. The participants in the interviews were purposively chosen to include employees from the engineering, project management, manufacturing, and quality assurance departments in order to obtain a broad insight into the ECM processes.

Semi structured interviews were directed by a set of queries to maintain consistency and elaboration of data collection and included questions about ECR work flow, decision making criteria, stakeholders, and technological tools in change management. The study setting offered a wealth of contextual information that helped to illuminate the nature of ECMs within a complex organizational environment. Another source of data collection was the examination of historical records and documentation of previous ECRs to compare interview results with historical information about change management practices and organizational experiences with previous ECRs. The use of this multimethod approach helped in data source triangulation, thereby increasing the reliability and validity of the study findings. During the data collection process, much attention was paid to ethical concerns such as anonymity, voluntary participation, and compliance with data protection rules.

IV. RESULTS

By conducting interviews, we were able to preserve a comprehensive comprehension of the firm within the framework of the analysis. A crucial element is to determine the available support alternatives for future ECRs by utilizing and examining the request history. Based on that foundation, we established objectives for our analysis to strategize the KDD project. **Table 2** displays the objectives and potential avenues for accomplishing the goals through qualitative content analysis. Moreover, the interviews have revealed the need to specifically uncover term occurrences, trends, and groups or subclasses in certain features of ECRs.

Table 2. Objectives of ECR attributes to be studied

objective	ECR attribute to be studied
proof the benefit of a keyword search for key issues in the change history	cause of change
	solution
	affected products/projects
	technical extent of problem
formation of groups or subclasses of ECRs	cause of change
	solution
	affected products/projects
	technical extent of problem

In order to evaluate our methodology, we chose a dataset consisting of 1,402 modification requests from three separate development projects. Each project represents an automobile product that is being produced on a large scale. The data was collected over a period of three years. This data collection contains, among other things, information regarding the source of the change, a potential solution to the challenge, the advantages of the solution, and any relevant comments. Based on the interviewee's statements, we chose certain criteria as outlined in **Table 2**. They suggested examining the similarities in the causes of change that necessitate an ECR. Each change request includes a field for free text, in which the engineer provides a description of the cause using several sentences.

To align text DM with numerical DM, it is necessary to see the data in a textual dataset, such as one found on the web, as a vast data base from which we can extract novel and previously unknown data (Hearst, 1999:3-4). DM, as defined by Hearst, is the process of (semi) inevitably uncovering trends and patterns within extensive databases, typically with the intention of facilitating decision-making. It does not involve the process of uncovering new pieces of information within inventory datasets. She considers 'corpus-based computational linguistics' to be akin to conventional numerical DM. In this approach, statistical calculations are performed on extensive text collections to uncover valuable language patterns, as illustrated in the left-hand column of **Table 3** and **Table 4**. In addition, she discusses the relationship between corpus-based computational linguistics and NLP. Computational linguistics enhances text and language analysis, but it does not provide any insights into the external world. Computational linguistics encompasses several tasks such as part-of-speech tagging, word sense disambiguation, and multilingual dictionary generation.

Table 3. Categorization of DM and Text DM Methods According to Hearst's Work in 1999

	Finding patterns	**Finding nuggets**	
		Novel	Non-novel
Textual data	Computational linguistics	Real TDM	Information retrieval
Non-textual data	Standard data mining	?	Database queries

Thus, our KDD approach includes a component dedicated to Text Mining (TM). DM and TM are (semi) automated processes used to identify patterns in existing databases [67]. The primary distinction between the two fields is in the inherent characteristics of the information being analyzed. DM bases on examining structured data, while Text Mining focuses on examining unstructured data. TM is a process that identifies and extracts valuable information from texts, including information that was previously unknown. It also organizes this information into categories. Text mining is a systematic procedure used to extract meaningful and noteworthy patterns in order to gain insights from textual data sources. Roughly 90% of the world's data is stored in an unstructured format [68]. Unstructured data is experiencing exponential growth in the 21st century [69].

Computational text analysis [70] has emerged as a captivating area of research with numerous applications in the realm of communication research. Implementing this strategy can be challenging due to the need for expertise in many methodologies, and the lack of availability of the necessary software in commonly used statistical programs. However, conducting Text Mining on such large scales can only be accomplished with technologies that facilitate the entire KDD process. We selected RapidMiner [71] for our investigation and proceeded to model the process using the following steps:

- Access the contents of an Excel file.
- Nominal to Text Data Conversion
- Choose Characteristics
- Substituting specific sections of the document
- Segment documents into tokens.
- Apply a length filter to the tokens.
- Remove Stopwords
- Stemming refers to the procedure of reducing words to their base or root form. Generate a word vector.

Table 4. Hearst's Publication Titled "Untangling Text Data Mining" from the Year 1999

	Finding trends or patterns	**Finding nuggets**	**Retrieval of non-novel information**
Non-textual (numerical) data	Standard data mining (KDD): This technique "separates signal from noise" and aids in decision making	?	Database queries
Textual data	Text data mining (corpus-based computational linguistics)	'Ore extraction' * Finding syntactic and lexical patterns in texts	Information retrieval * Information that has previously been discovered, at

		* Real-time TDM automatic subcategorization data acquisition * Using text information to provide context for the outside world: compares category assignment distributions to find novel patterns * Finds the emergence of fresh topics in text libraries * Exploratory data analysis that finds connections and generates new theories by collaborating between a human researcher and text-mining technologies	least by the requisite papers' authors * Advanced IR: "Term associations" are automatically generated to help with query expansion * Using co-citation analysis in a text collection to identify overarching themes * Using text clustering in a text corpus to produce theme overviews * Text classification using a predetermined set of labels * Summarisation * Web search

The following processes are typically used in Text Mining (TM) applications [72] and can be enhanced by additional methods such as classification and regression, clustering and segmentation, and correlation and dependency computing. For instance, in the field of clustering, numerous jobs in text mining require the arrangement of text into groups where documents within the same group exhibit similarity, while documents from separate groups do not [73]. Clustering is the term used to describe the grouping process. The main objective of text clustering is to either arrange information in a manner that simplifies retrieval and search or to implement an automated classification of documents. Text clustering has been performed in various domains. In particular, it has been used to classify crime features, including the type of crime, the location of the crime, and the types of weapons mentioned in crime reports [74]. Moreover, it has been applied to the improvement of the legal practice areas` taxonomy, which facilitates classification and categorization [75]. Text clustering has been applied in classifying the results of search query and establishing the taxonomy of documents in order to enhance document retrieval systems, and web-based search engines [76]. Before carrying out text clustering, researchers have to describe the distances between texts such as Euclidean distance. This distance could be computed on the initial variables or on the reduced variables, for instance after dimensionality reduction.

Clustering approaches can be grouped into two major clusters: partitional and hierarchical [77]. Hierarchical clustering algorithms can be categorized into two types: agglomerative and divisive. Agglomerative algorithms first treat every object as a distinct cluster, gradually merging clusters until every object is a part of a single cluster. On the other hand, divisive algorithms begin by grouping all items into one cluster and then recursively subdivide clusters until every item creates into separate group. The process of combining (or dividing) groups is illustrated by dendrogram or a tree [78]. In partitional clustering, uses have to define the number of groups beforehand. Clusters are then constructed by maximizing an objective function, typically based on the distances between items and the centers of the groups they are assigned to. The widely-used k-means technique is an instance of partitional clustering, as stated by Xiao and Yu [79]. An essential obstacle in clustering is the selection of the optimal number of clusters to create. As clustering is an investigative technique, a typical approach is to test various numbers of groups and utilize cluster valuation measures to make a decision. Two examples of quality metrics are the silhouette coefficient and the Dunn index [80].

Typical text cleaning techniques [81] involve removing unnecessary characters such as superfluous whitespaces and formatting tags. It also includes segmenting the text, converting it to lowercase, removing stop words, and performing word stemming. In our experience, it can be beneficial to conduct a spell-check on open-ended surveys and other informally-generated texts, like personal emails or SMS texts, to rectify any misspelled words. In the case of web pages, it is essentially to eliminate HTML or XML elements as they do not contribute any significant material. Consequently, the final outcome is text data that has been deprived of all insignificant words and characters. Text segmentation, as described by Beeferman, Berger, and Lafferty [82], refers to the procedure of splitting text into individual phrases and words. Stop words, like prepositions (such as “for”, “and”, “of”, and “the”) and conjunctions, are words with minimal data value that do not significantly contribute to the definition of the prepositions. The process of “stemming” standardizes the depiction of words that are semantically similar, such as replacing the phrases “ensuring,” “ensures,” and “ensured” with “ensure”. As these strategies remove words, they also help decrease the quantity of the vocabulary. We performed the process and generated a word vector by calculating term occurrences for all change requests, based on the attribute cause of change. A total of 5.783 distinct terms were identified in the dataset, allowing for analysis from various angles. Almost every ECR record contains information on the parts that have been damaged. At this higher level of analysis, the outcomes are to be expected. However, they provide the opportunity to conduct thorough and precise assessments.

The KDD process represents a continuous loop from transformation, data selection, data mining, preprocessing, and result interpretation. It clearly shows the need to incorporate business understanding as a part of the first step in the KDD procedure while ensuring that the process is optimally oriented towards the goals of the business and its ability to gain valuable insights from the data. In **Fig 2** it presents the frequency distribution of term occurrences for the entire ECR dataset classified into different clusters according to the keywords used. It points out the frequency of each term to help visualize which areas are most addressed by the current ECM and may focus on in the future to improve these processes.

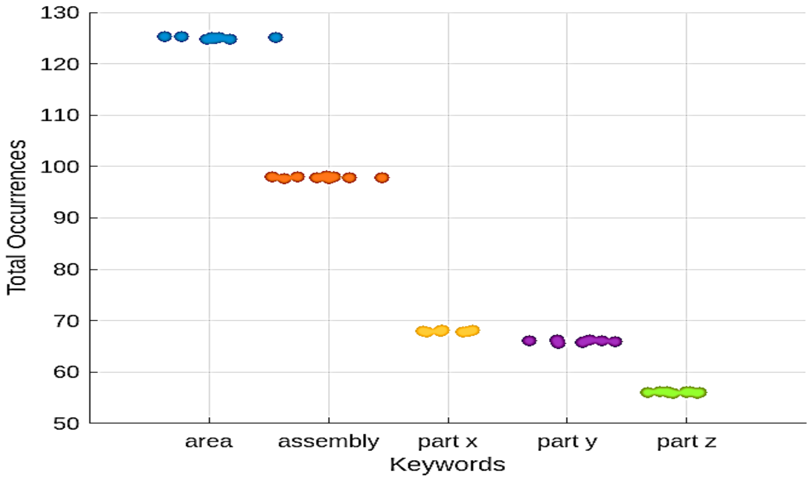

Fig 2. Total Occurrences with Clusters.

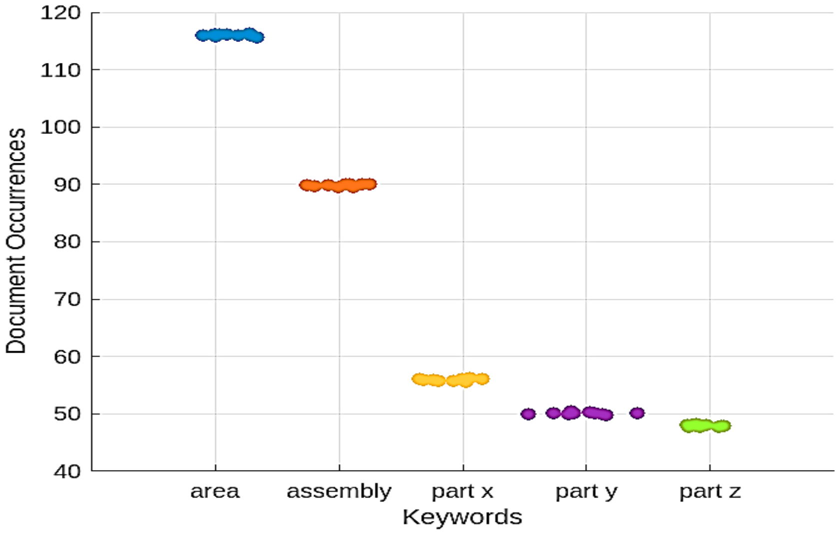

Fig 3. Document Occurrences with Clusters.

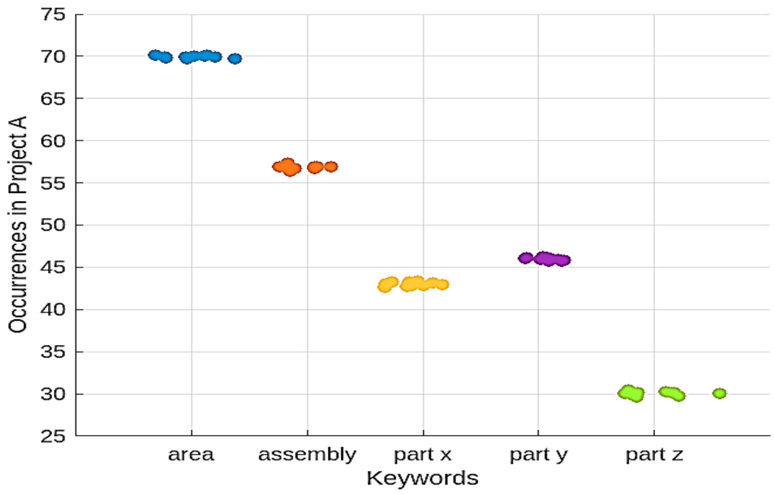

Fig 4. Project a Occurrences with Clusters.

Fig 3 shows the identified terms in the individual documents and how they are grouped into clusters. This useful for understanding how often certain terms are used in various documents to gain an idea of trends and potential problems that may occur during the engineering change process. In **Fig 4**, the focus is made on the term frequency of Project A only, which outlines the frequency of the use of every keyword related to this particular project. It enables one to pick certain rhythms and beats in a project and address them or change them individually with the aim of enhancing productivity.

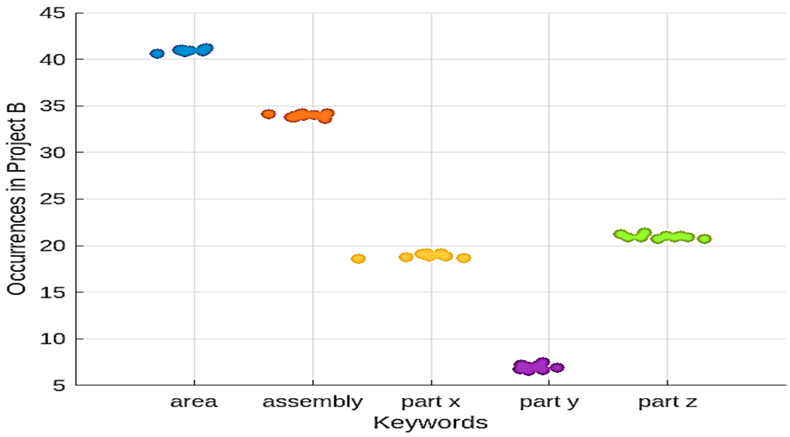

Fig 5. Project B Occurrences with Clusters.

The term occurrence for Project B is presented in **Fig 5** categorized into clusters. This type of mapping comes in handy when defining the precise issues and corresponding terms of Project B so as to enable accurate management of the challenges that may be expected in the course of the project. The distribution of the term occurrences for Project C grouped by clusters is shown in **Fig 6** to determine the current terms and trends in Project C, which is useful for developing Project C-specific approaches and improving the ECM practices. The allocation of keywords serves as an initial step for further computations and visual representations. **Fig 7** displays a total of 1,403 records and examines the frequency of the terms "areas" and "assembly," as well as the occurrences of "part x," "part y," and "part z." It is evident that 7 requests contain both keywords and originate from two distinct projects (shown by the color of the dots). The following essential phases involve conducting further Text Mining computations and subsequently interpreting and evaluating our findings. The intention is to expand the data set and analyze practically all existing records of ECRs.

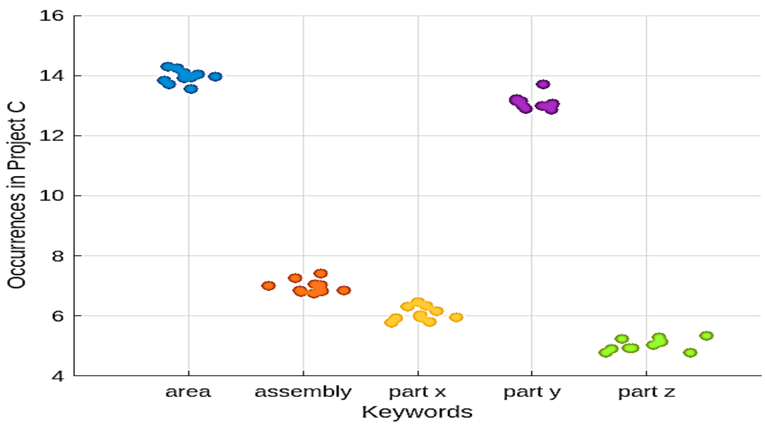

Fig 6. Project C Occurrences with Clusters.

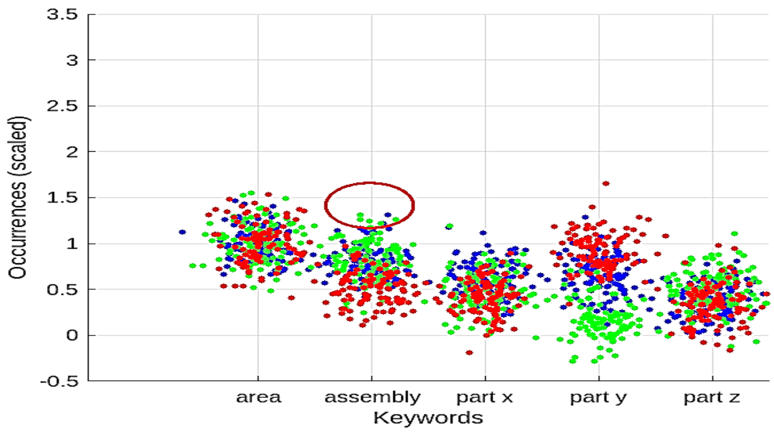

Fig 7. Scatter of Keyword Occurrences.

V. DISCUSSION

The formal implementation of the engineering change process typically begins in many firms once a design project has achieved its initial baseline, such as the "System Design Review" [83]. At this point, all project documentation is submitted for approval and subsequently restricted from further modifications. The baseline serves as a reference point prior to commencing new activities that require precise information, such as detailed design or production. In certain sectors, like as the defense industry, the customer actively participates in reviewing and approving the progress made up to a certain point, known as the baseline. Every document undergoes verification in accordance with the specifications and is then approved for publication. After being released, documents can only be modified through the implementation of a proper EC process. The EC process is an integral component of the broader configuration management (CM) process [84]. Any alteration to the product or its associated documents results in a modification of the product conformation. Consequently, each time a fresh iteration of a product document is generated, it is imperative to associate it with the appropriate persons responsible for its production. When a defective product is reported to the company, it is essential to be able to locate the associated document modifications with ease. This is especially crucial in the defense industry, particularly when multiple product variations are being manufactured [85]. Therefore, the establishment of a proficient and productive EC process is vital for the success of a CM process.

Through the analysis of the interviews, we have assessed the present state of Engineering Change Management (ECM). Research has shown that the implementation of ECM (Electronic Control Modules) can effectively mitigate the adverse effects of ECs (Electronic Cigarettes) [86]. In [87], ECM stands for enterprise change management, which encompasses the organization, control, and implementation of engineering changes over the full product lifecycle, from concept assortment to production wind-down and support [88]. The objectives of ECM are to minimize the quantity of exceptional circumstances (ECs), to efficiently choose ECs when they arise, to properly carry out ECs, and to consistently get knowledge from the implementation process [89]. In order to achieve these aims, a range of ECM strategies and technologies have been devised. A number of case studies and surveys have examined the prevalence of ECM techniques and tools in various industries [90, 91, 92]. Van De Ven and Huber [93] stated that future ECM is expected to vary based on relative factors such as manufacturing volume, product customization level, uncertainty degree, and intrinsic product intricacy. Nevertheless, the existing research fail to clearly differentiate the contextual circumstances of various production contexts that could impact the utilization of these methods, techniques, and technologies. Therefore, it remains uncertain if ECM strategies and tools can be equally effective in various production situations.

Prior to initiating any KDD initiatives, it is imperative to do the Business Understanding phase. We were able to determine the goals for our KDD methodology. Having a clear comprehension of the business allows data scientists to pinpoint the particular data sources that are highly pertinent to the organization's goals. By evaluating the quality, comprehensiveness, and correctness of the data at hand, individuals can make well-informed choices about data preprocessing, cleansing, and augmentation methods in order to guarantee dependable outcomes. According to Mendonça, Pereira, and Godinho [94], knowledge is considered to be the most intangible and immeasurable resource for organizations, corporations, and individuals. The success of companies, educational institutes, and firms depends on the quality of knowledge they possess. Therefore, organizations must effectively generate, find, and distribute both implicit and explicit knowledge within the organization. They must also preserve their expertise to stay competitive in the business world [95].

Knowledge management is a form of structural activity that is dependent on context and varies in its application. Sikes [96] outlined the five essential processes that define the information Management process of the APO. These steps include recognizing, creating, storing, sharing, and using information. The process of identification and creation are undeniably equivalent to knowledge discovery, and they serve as the foundation for any effective knowledge management process in any organization [97]. Bollinger and Smith [98] emphasized that Knowledge management is the study of how to effectively utilize and enhance an organization's knowledge assets in order to reinforce its objectives. The crucial aspect of knowledge discovery processes is the development of knowledge systems, which can be enhanced by utilizing data mining technologies. These technologies are particularly useful for identifying new connections among the available data. The resulting model can then be used to forecast, classify, and apply information as valuable assets in business intelligence [99]. Consequently, there was a consistent occurrence of ambiguous interpretations, thus necessitating a clear comprehension of data mining and KDD.

During the second phase, the Text Mining method produced word vectors. Through this process, we have successfully found pertinent keywords that allow us to discern patterns within the data. Word vectorization is a technique that involves assigning numerical values to terms in a lexicon. These vectors are applicable in multiple NLP ML models to perform tasks such as Text Similarity, Topic modeling, POS detection, and prediction. Text mining is an emerging field in computer science that has close ties to knowledge management, NLP, data retrieval, ML, and data mining. Text mining aims to get valuable insights from unstructured textual material by identifying and analyzing intriguing patterns [100]. Yan et al. [101] examine various methods for identifying global patterns in text using the "bag-of-words" (BOW) model outlined in [102]. The methods discussed include automated classification, clustering, and dimensionality reduction. The demonstration of pattern discovery will be exemplified by two applications derived from our recent research: the display of search engine results from a Web meta-search engine [103], and the conception of co-authorship connections inevitably derived from a semi-structured collection of papers that describe scholars in [104]. We utilized two potential methodologies. First, we employed findings to determine the specific circumstances that prompt change requests. Secondly, we employed data as a foundation to create further support for individuals applying for change requests. RapidMiner enables the modeling of the Text Mining process with the help of available requests [105]. It consists of a variable of operators that allows constructing sufficient analysis procedure, by means of which necessary data can be viewed from different angles.

VI. CONCLUSION AND OUTLOOK

Historical ECR data contributes to the improvement of the ECM. In our study, using a real context of a global automobile firm, we identified that there are patterns and information in ECR data, which are still rather unexplored. The interviews with the main stakeholders provided an understanding of the significance of integrating historical data into the decision-making process. From the 3 development projects namely Project A, B, and C, 1,402 ECRs were analyzed and meaningful term frequency, pattern and subclasses of ECR attributes, as well as cause of change, solution, affected products/projects and technical extent of the problem were determined. We concluded that a keyword search can be effective for the identification of the key issues in the change history. Thus, the use of KDD and Text Mining methods enabled the distinction of the most frequent patterns and groups or subclasses of ECRs. The studies carried out established that the integration of KDD and Text Mining into ECM results in enhanced management of the ECRs, decreased development cycles, and controlled costs. The repeated cycle of research and improvements helped to make our study solid and pragmatic and highlighted the importance of the use of historical data in improving the current situations and developing new strategies and solutions in highly technical contexts. Thus, this research highlights the prospects of data science solutions in enhancing ECM practices to be more efficient, effective, and responsive to organizational goals.

CRediT Author Statement

The authors confirm contribution to the paper as follows:
Conceptualization: Zheng Xinyu and He Xin; **Methodology:** Zheng Xinyu and He Xin; **Supervision:** Zheng Xinyu; **Validation:** He Xin; **Writing- Reviewing and Editing:** Zheng Xinyu and He Xin. All authors reviewed the results and approved the final version of the manuscript.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests

References

[1]. W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus, "Knowledge discovery in databases: an overview," ˜ the œAI Magazine/AI Magazine, vol. 13, no. 3, pp. 57–70, Sep. 1992, doi: 10.1609/aimag.v13i3.1011.
[2]. U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," ˜ the œAI Magazine/AI Magazine, vol. 17, no. 3, pp. 37–54, Mar. 1996, doi: 10.1609/aimag.v17i3.1230.
[3]. Chen, Chiang, and Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Quarterly, vol. 36, no. 4, p. 1165, 2012, doi: 10.2307/41703503.
[4]. B. Molina-Coronado, U. Mori, A. Mendiburu, and J. Miguel-Alonso, "Survey of Network Intrusion Detection Methods From the Perspective of the Knowledge Discovery in Databases Process," IEEE Transactions on Network and Service Management, vol. 17, no. 4, pp. 2451–2479, Dec. 2020, doi: 10.1109/tnsm.2020.3016246.

[5]. M. Pejić Bach, Ž. Krstić, S. Seljan, and L. Turulja, "Text Mining for Big Data Analysis in Financial Sector: A Literature Review," Sustainability, vol. 11, no. 5, p. 1277, Feb. 2019, doi: 10.3390/su11051277.
[6]. L. ZINGALES, "Presidential Address: Does Finance Benefit Society?," The Journal of Finance, vol. 70, no. 4, pp. 1327–1363, Jul. 2015, doi: 10.1111/jofi.12295.
[7]. A. Wasmer, G. Staub, and R. W. Vroom, "An industry approach to shared, cross-organisational engineering change handling - The road towards standards for product data processing," Computer-Aided Design, vol. 43, no. 5, pp. 533–545, May 2011, doi: 10.1016/j.cad.2010.10.002.
[8]. P. Deb and E. C. Norton, "Modeling Health Care Expenditures and Use," Annual Review of Public Health, vol. 39, no. 1, pp. 489–505, Apr. 2018, doi: 10.1146/annurev-publhealth-040617-013517.
[9]. B. Succar, "Building information modelling framework: A research and delivery foundation for industry stakeholders," Automation in Construction, vol. 18, no. 3, pp. 357–375, May 2009, doi: 10.1016/j.autcon.2008.10.003.
[10]. P. Hennebert, H. A. van der Sloot, F. Rebischung, R. Weltens, L. Geerts, and O. Hjelmar, "Hazard property classification of waste according to the recent propositions of the EC using different methods," Waste Management, vol. 34, no. 10, pp. 1739–1751, Oct. 2014, doi: 10.1016/j.wasman.2014.05.021.
[11]. T. A. W. Jarratt, C. M. Eckert, N. H. M. Caldwell, and P. J. Clarkson, "Engineering change: an overview and perspective on the literature," Research in Engineering Design, vol. 22, no. 2, pp. 103–124, Dec. 2010, doi: 10.1007/s00163-010-0097-y.
[12]. L. W. Busenitz and J. B. Barney, "Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making," Journal of Business Venturing, vol. 12, no. 1, pp. 9–30, Jan. 1997, doi: 10.1016/s0883-9026(96)00003-1.
[13]. O. Kolomiyets and M.-F. Moens, "A survey on question answering technology from an information retrieval perspective," Information Sciences, vol. 181, no. 24, pp. 5412–5434, Dec. 2011, doi: 10.1016/j.ins.2011.07.047.
[14]. B. Nadia, G. Gregory, and T. Vince, "Engineering change request management in a new product development process," European Journal of Innovation Management, vol. 9, no. 1, pp. 5–19, Jan. 2006, doi: 10.1108/14601060610639999.
[15]. H. Gmelin and S. Seuring, "Achieving sustainable new product development by integrating product life-cycle management capabilities," International Journal of Production Economics, vol. 154, pp. 166–177, Aug. 2014, doi: 10.1016/j.ijpe.2014.04.023.
[16]. T. L. Saaty, "An Exposition of the AHP in Reply to the Paper 'Remarks on the Analytic Hierarchy Process,'" Management Science, vol. 36, no. 3, pp. 259–268, Mar. 1990, doi: 10.1287/mnsc.36.3.259.
[17]. C. Rowe, J. G. Birnberg, and M. D. Shields, "Effects of organizational process change on responsibility accounting and managers' revelations of private knowledge," Accounting, Organizations and Society, vol. 33, no. 2–3, pp. 164–198, Feb. 2008, doi: 10.1016/j.aos.2006.12.002.
[18]. T. Zimmermann, A. Zeller, P. Weissgerber, and S. Diehl, "Mining version histories to guide software changes," IEEE Transactions on Software Engineering, vol. 31, no. 6, pp. 429–445, Jun. 2005, doi: 10.1109/tse.2005.72.
[19]. K. N. Otto and K. L. Wood, "Product Evolution: A Reverse Engineering and Redesign Methodology," Research in Engineering Design, vol. 10, no. 4, pp. 226–243, Dec. 1998, doi: 10.1007/s001639870003.
[20]. N. Bertrand, J. Wu, X. Xu, N. Kamaly, and O. C. Farokhzad, "Cancer nanotechnology: The impact of passive and active targeting in the era of modern cancer biology," Advanced Drug Delivery Reviews, vol. 66, pp. 2–25, Feb. 2014, doi: 10.1016/j.addr.2013.11.009.
[21]. A. Usai, M. Pironti, M. Mital, and C. Aouina Mejri, "Knowledge discovery out of text data: a systematic review via text mining," Journal of Knowledge Management, vol. 22, no. 7, pp. 1471–1488, May 2018, doi: 10.1108/jkm-11-2017-0517.
[22]. P. Ristoski and H. Paulheim, "Semantic Web in data mining and knowledge discovery: A comprehensive survey," Journal of Web Semantics, vol. 36, pp. 1–22, Jan. 2016, doi: 10.1016/j.websem.2016.01.001.
[23]. J. Zhang, "Knowledge Learning With Crowdsourcing: A Brief Review and Systematic Perspective," IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 5, pp. 749–762, May 2022, doi: 10.1109/jas.2022.105434.
[24]. M. Mezzanzanica, R. Boselli, M. Cesarini, and F. Mercorio, "A model-based evaluation of data quality activities in KDD," Information Processing & Management, vol. 51, no. 2, pp. 144–166, Mar. 2015, doi: 10.1016/j.ipm.2014.07.007.
[25]. G. G. Chowdhury, "Natural language processing," Annual Review of Information Science and Technology, vol. 37, no. 1, pp. 51–89, Jan. 2003, doi: 10.1002/aris.1440370103.
[26]. L. A. KURGAN and P. MUSILEK, "A survey of Knowledge Discovery and Data Mining process models," The Knowledge Engineering Review, vol. 21, no. 1, pp. 1–24, Mar. 2006, doi: 10.1017/s0269888906000737.
[27]. A. Rotondo and F. Quilligan, "Evolution Paths for Knowledge Discovery and Data Mining Process Models," SN Computer Science, vol. 1, no. 2, Mar. 2020, doi: 10.1007/s42979-020-0117-6.
[28]. W.-T. Wu et al., "Data mining in clinical big data: the frequently used databases, steps, and methodological models," Military Medical Research, vol. 8, no. 1, Aug. 2021, doi: 10.1186/s40779-021-00338-z.
[29]. V. Plotnikova, M. Dumas, and F. P. Milani, "Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements," Data & Knowledge Engineering, vol. 139, p. 102013, May 2022, doi: 10.1016/j.datak.2022.102013.
[30]. H. Wiemer, L. Drowatzky, and S. Ihlenfeldt, "Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model," Applied Sciences, vol. 9, no. 12, p. 2407, Jun. 2019, doi: 10.3390/app9122407.
[31]. S. Studer et al., "Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology," Machine Learning and Knowledge Extraction, vol. 3, no. 2, pp. 392–413, Apr. 2021, doi: 10.3390/make3020020.
[32]. G. Schuh, J.-P. Prote, M. Luckert, F. Basse, V. Thomson, and W. Mazurek, "Adaptive Design of Engineering Change Management in Highly Iterative Product Development," Procedia CIRP, vol. 70, pp. 72–77, 2018, doi: 10.1016/j.procir.2018.02.016.
[33]. A. G. Riess et al., "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant," The Astronomical Journal, vol. 116, no. 3, pp. 1009–1038, Sep. 1998, doi: 10.1086/300499.
[34]. S.-L. Chou, Y. Pan, J.-Z. Wang, H.-K. Liu, and S.-X. Dou, "Small things make a big difference: binder effects on the performance of Li and Na batteries," Physical Chemistry Chemical Physics, vol. 16, no. 38, pp. 20347–20359, 2014, doi: 10.1039/c4cp02475c.
[35]. R. Jervis 1, "Reports, politics, and intelligence failures: The case of Iraq," Journal of Strategic Studies, vol. 29, no. 1, pp. 3–52, Feb. 2006, doi: 10.1080/01402390600566282.
[36]. D. S. Marks et al., "Protein 3D Structure Computed from Evolutionary Sequence Variation," PLoS ONE, vol. 6, no. 12, p. e28766, Dec. 2011, doi: 10.1371/journal.pone.0028766.
[37]. K. M. Eisenhardt and J. A. Martin, "Dynamic capabilities: what are they?," Strategic Management Journal, vol. 21, no. 10–11, pp. 1105–1121, 2000, doi: 10.1002/1097-0266(200010/11)21:10/11<1105::aid-smj133>3.0.co;2-e.
[38]. D. Beran et al., "Research capacity building—obligations for global health partners," The Lancet Global Health, vol. 5, no. 6, pp. e567–e568, Jun. 2017, doi: 10.1016/s2214-109x(17)30180-8.
[39]. T. Bekhuis, "Conceptual biology, hypothesis discovery, and text mining: Swanson's legacy," Biomedical Digital Libraries, vol. 3, no. 1, Apr. 2006, doi: 10.1186/1742-5581-3-2.
[40]. P. Carrillo, J. Harding, and A. Choudhary, "Knowledge discovery from post-project reviews," Construction Management and Economics, vol. 29, no. 7, pp. 713–723, Jul. 2011, doi: 10.1080/01446193.2011.588953.
[41]. A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," Expert Systems with Applications, vol. 41, no. 4, pp. 1432–1462, Mar. 2014, doi: 10.1016/j.eswa.2013.08.042.

[42]. R. Zheng, J. Li, H. Chen, and Z. Huang, “A framework for authorship identification of online messages: Writing-style features and classification techniques,” Journal of the American Society for Information Science and Technology, vol. 57, no. 3, pp. 378–393, Dec. 2005, doi: 10.1002/asi.20316.

[43]. B. Haddow, R. Bawden, A. V. M. Barone, J. Helcl, and A. Birch, “Survey of Low-Resource Machine Translation,” Computational Linguistics, vol. 48, no. 3, pp. 673–732, 2022, doi: 10.1162/coli_a_00446.

[44]. C. Dunne, B. Shneiderman, R. Gove, J. Klavans, and B. Dorr, “Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization,” Journal of the American Society for Information Science and Technology, vol. 63, no. 12, pp. 2351–2369, Nov. 2012, doi: 10.1002/asi.22652.

[45]. C. Bizer, T. Heath, and T. Berners-Lee, “Linked Data - The Story So Far,” International Journal on Semantic Web and Information Systems, vol. 5, no. 3, pp. 1–22, Jul. 2009, doi: 10.4018/jswis.2009081901.

[46]. B. Abu-Salih, “Domain-specific knowledge graphs: A survey,” Journal of Network and Computer Applications, vol. 185, p. 103076, Jul. 2021, doi: 10.1016/j.jnca.2021.103076.

[47]. L. Soibelman and H. Kim, “Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases,” Journal of Computing in Civil Engineering, vol. 16, no. 1, pp. 39–48, Jan. 2002, doi: 10.1061/(asce)0887-3801(2002)16:1(39).

[48]. A. García Rudolph, “Supporting the design of sequences of cumulative activities impacting on multiple areas through a data mining approach : application to design of cognitive rehabilitation programs for traumatic brain injury patients”, doi: 10.5821/dissertation-2117-96241.

[49]. Z. Bosnjak, O. Grljevic, and S. Bosnjak, “CRISP-DM as a framework for discovering knowledge in small and medium sized enterprises’ data,” 2009 5th International Symposium on Applied Computational Intelligence and Informatics, pp. 509–514, May 2009, doi: 10.1109/saci.2009.5136302.

[50]. M. Riesener, C. Doelle, M. Mendl-Heinisch, and G. Schuh, “Literature Based Derivation of a Framework to Evaluate Engineering Change Requests,” 2019 Portland International Conference on Management of Engineering and Technology (PICMET), pp. 1–6, Aug. 2019, doi: 10.23919/picmet.2019.8893906.

[51]. S. Y. Rhee, J. Dickerson, and D. Xu, “BIOINFORMATICS AND ITS APPLICATIONS IN PLANT BIOLOGY,” Annual Review of Plant Biology, vol. 57, no. 1, pp. 335–360, Jun. 2006, doi: 10.1146/annurev.arplant.56.032604.144103.

[52]. H. Etzkowitz, A. Webster, C. Gebhardt, and B. R. C. Terra, “The future of the university and the university of the future: evolution of ivory tower to entrepreneurial paradigm,” Research Policy, vol. 29, no. 2, pp. 313–330, Feb. 2000, doi: 10.1016/s0048-7333(99)00069-4.

[53]. D. Perrotta, J. Faria, M. Araajo, A. Tereso, and G. Fernandes, “Project change request: A proposal for managing change in industrialization projects,” 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 1525–1529, Dec. 2017, doi: 10.1109/ieem.2017.8290148.

[54]. K. Pavitt, “Sectoral patterns of technical change: Towards a taxonomy and a theory,” Research Policy, vol. 13, no. 6, pp. 343–373, Dec. 1984, doi: 10.1016/0048-7333(84)90018-0.

[55]. L. Jokinen, V. Vainio, and A. Pulkkinen, “Engineering Change Management Data Analysis from the Perspective of Information Quality,” Procedia Manufacturing, vol. 11, pp. 1626–1633, 2017, doi: 10.1016/j.promfg.2017.07.312.

[56]. J. Han, S.-H. Lee, and P. Nyamsuren, “An integrated engineering change management process model for a project-based manufacturing,” International Journal of Computer Integrated Manufacturing, vol. 28, no. 7, pp. 745–752, Jul. 2014, doi: 10.1080/0951192x.2014.924342.

[57]. M. E. Sharp, T. D. Hedberg, W. Z. Bernstein, and S. Kwon, “Feasibility study for an automated engineering change process,” International Journal of Production Research, vol. 59, no. 16, pp. 4995–5010, Mar. 2021, doi: 10.1080/00207543.2021.1893900.

[58]. N. Do, “Integration of engineering change objects in product data management databases to support engineering change analysis,” Computers in Industry, vol. 73, pp. 69–81, Oct. 2015, doi: 10.1016/j.compind.2015.08.002.

[59]. G. Q. Huang, W. Y. Yee, and K. L. Mak, “Development of a web-based system for engineering change management,” Robotics and Computer-Integrated Manufacturing, vol. 17, no. 3, pp. 255–267, Jun. 2001, doi: 10.1016/s0736-5845(00)00058-2.

[60]. T.-K. Peng and A. J. C. Trappey, “A step toward STEP-compatible engineering data management: the data models of product structure and engineering changes,” Robotics and Computer-Integrated Manufacturing, vol. 14, no. 2, pp. 89–109, Apr. 1998, doi: 10.1016/s0736-5845(97)00016-1.

[61]. K. Rouibah and K. R. Caskey, “Change management in concurrent engineering from a parameter perspective,” Computers in Industry, vol. 50, no. 1, pp. 15–34, Jan. 2003, doi: 10.1016/s0166-3615(02)00138-0.

[62]. U. Lindemann, E. F. Shakirov, N. Kattner, C. Fortin, and I. K. Uzhinsk, “Reducing the uncertainty in engineering change management using historical data and simulation modelling: a process twin concept,” International Journal of Product Lifecycle Management, vol. 13, no. 1, p. 1, Jan. 2021, doi: 10.1504/ijplm.2021.10037271.

[63]. G. S. Becker, “Crime and Punishment: An Economic Approach,” Journal of Political Economy, vol. 76, no. 2, pp. 169–217, Mar. 1968, doi: 10.1086/259394.

[64]. R. B. Cooper and R. W. Zmud, “Information Technology Implementation Research: A Technological Diffusion Approach,” Management Science, vol. 36, no. 2, pp. 123–139, Feb. 1990, doi: 10.1287/mnsc.36.2.123.

[65]. R. Gangl, T. Gollmann, and T. Gruchmann, “From Engineering Change to Enterprise Change Management: An empirical study on CM2 processes in the automotive industry,” Procedia CIRP, vol. 122, pp. 629–634, 2024, doi: 10.1016/j.procir.2024.01.090.

[66]. M. Callon and V. Rabeharisoa, “Research ‘in the wild’ and the shaping of new social identities,” Technology in Society, vol. 25, no. 2, pp. 193–204, Apr. 2003, doi: 10.1016/s0160-791x(03)00021-6.

[67]. G. George, E. C. Osinga, D. Lavie, and B. A. Scott, “Big Data and Data Science Methods for Management Research,” Academy of Management Journal, vol. 59, no. 5, pp. 1493–1507, Oct. 2016, doi: 10.5465/amj.2016.4005.

[68]. D. Delen and M. D. Crossland, “Seeding the survey and analysis of research literature with text mining,” Expert Systems with Applications, vol. 34, no. 3, pp. 1707–1720, Apr. 2008, doi: 10.1016/j.eswa.2007.01.035.

[69]. J. Sheng, J. Amankwah-Amoah, and X. Wang, “Technology in the 21st century: New challenges and opportunities,” Technological Forecasting and Social Change, vol. 143, pp. 321–335, Jun. 2019, doi: 10.1016/j.techfore.2018.06.009.

[70]. P. DiMaggio, “Adapting computational text analysis to social science (and vice versa),” Big Data & Society, vol. 2, no. 2, Dec. 2015, doi: 10.1177/2053951715602908.

[71]. P. Ristoski, C. Bizer, and H. Paulheim, “Mining the Web of Linked Data with RapidMiner,” Journal of Web Semantics, vol. 35, pp. 142–151, Dec. 2015, doi: 10.1016/j.websem.2015.06.004.

[72]. A. Abbe, C. Grouin, P. Zweigenbaum, and B. Falissard, “Text mining applications in psychiatry: a systematic literature review,” International Journal of Methods in Psychiatric Research, vol. 25, no. 2, pp. 86–100, Jul. 2015, doi: 10.1002/mpr.1481.

[73]. Y.-H. Tseng, C.-J. Lin, and Y.-I. Lin, “Text mining techniques for patent analysis,” Information Processing & Management, vol. 43, no. 5, pp. 1216–1247, Sep. 2007, doi: 10.1016/j.ipm.2006.11.011.

[74]. G. Bello-Orgaz, J. J. Jung, and D. Camacho, “Social big data: Recent achievements and new challenges,” Information Fusion, vol. 28, pp. 45–59, Mar. 2016, doi: 10.1016/j.inffus.2015.08.005.

[75]. A. Boonstra and M. Broekhuis, “Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions,” BMC Health Services Research, vol. 10, no. 1, Aug. 2010, doi: 10.1186/1472-6963-10-231.

[76]. S. Raghavan, "Digital forensic research: current state of the art," CSI Transactions on ICT, vol. 1, no. 1, pp. 91–114, Nov. 2012, doi: 10.1007/s40012-012-0008-7.

[77]. F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery/Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, vol. 2, no. 1, pp. 86–97, Dec. 2011, doi: 10.1002/widm.53.

[78]. M. Mouchet, F. Guilhaumon, S. Villéger, N. W. H. Mason, J. Tomasini, and D. Mouillot, "Towards a consensus for calculating dendrogram-based functional diversity indices," Oikos, vol. 117, no. 5, pp. 794–800, Apr. 2008, doi: 10.1111/j.0030-1299.2008.16594.x.

[79]. Y. Xiao and J. Yu, "Partitive clustering (K-means family)," Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery/Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, vol. 2, no. 3, pp. 209–225, Mar. 2012, doi: 10.1002/widm.1049.

[80]. W. G. McMillan and J. E. Mayer, "The Statistical Thermodynamics of Multicomponent Systems," The Journal of Chemical Physics, vol. 13, no. 7, pp. 276–305, Jul. 1945, doi: 10.1063/1.1724036.

[81]. H. H. Malik and V. S. Bhardwaj, "Automatic Training Data Cleaning for Text Classification," 2011 IEEE 11th International Conference on Data Mining Workshops, pp. 442–449, Dec. 2011, doi: 10.1109/icdmw.2011.36.

[82]. D. Beeferman, A. Berger, and J. Lafferty, "Statistical Models for Text Segmentation," Machine Learning, vol. 34, no. 1–3, pp. 177–210, Feb. 1999, doi: 10.1023/a:1007506220214.

[83]. C. J. Leising, B. Sherwood, M. Adler, R. R. Wessen, and F. M. Naderi, "Recent improvements in JPL's mission formulation process," 2010 IEEE Aerospace Conference, pp. 1–12, Mar. 2010, doi: 10.1109/aero.2010.5446881.

[84]. R. Jain, A. Chandrasekaran, and O. Erol, "A framework for end-to-end approach to Systems Integration," International Journal of Industrial and Systems Engineering, vol. 5, no. 1, p. 79, 2010, doi: 10.1504/ijise.2010.029763.

[85]. M. Hobday, "Product complexity, innovation and industrial organisation," Research Policy, vol. 26, no. 6, pp. 689–710, Feb. 1998, doi: 10.1016/s0048-7333(97)00044-9.

[86]. M. Chandy et al., "Addressing Cardiovascular Toxicity Risk of Electronic Nicotine Delivery Systems in the Twenty-First Century: 'What Are the Tools Needed for the Job?' and 'Do We Have Them?,'" Cardiovascular Toxicology, vol. 24, no. 5, pp. 435–471, Mar. 2024, doi: 10.1007/s12012-024-09850-9.

[87]. N. Iakymenko, A. Romsdal, E. Alfnes, M. Semini, and J. O. Strandhagen, "Status of engineering change management in the engineer-to-order production environment: insights from a multiple case study," International Journal of Production Research, vol. 58, no. 15, pp. 4506–4528, May 2020, doi: 10.1080/00207543.2020.1759836.

[88]. B. Hamraz, N. H. M. Caldwell, and P. J. Clarkson, "A Holistic Categorization Framework for Literature on Engineering Change Management," Systems Engineering, vol. 16, no. 4, pp. 473–505, Aug. 2013, doi: 10.1002/sys.21244.

[89]. J. Wilberg, F. Elezi, I. D. Tommelein, and U. Lindemann, "Using a Systemic Perspective to Support Engineering Change Management," Procedia Computer Science, vol. 61, pp. 287–292, 2015, doi: 10.1016/j.procs.2015.09.217.

[90]. J. vom Brocke, A. Simons, and A. Cleven, "Towards a business process-oriented approach to enterprise content management: the ECM-blueprinting framework," Information Systems and e-Business Management, vol. 9, no. 4, pp. 475–496, Jan. 2010, doi: 10.1007/s10257-009-0124-6.

[91]. D. T. Abdurrahaman, A. Owusu, and A. S. Bakare, "Evaluating Factors Affecting User Satisfaction in University Enterprise Content Management (ECM) Systems," Electronic Journal of Information Systems Evaluation, vol. 23, no. 1, Feb. 2020, doi: 10.34190/ejise.20.23.1.001.

[92]. J. A. Alalwan, M. A. Thomas, and H. R. Weistroffer, "Decision support capabilities of enterprise content management systems: An empirical investigation," Decision Support Systems, vol. 68, pp. 39–48, Dec. 2014, doi: 10.1016/j.dss.2014.09.002.

[93]. A. H. van de Ven and G. P. Huber, "Longitudinal Field Research Methods for Studying Processes of Organizational Change," Organization Science, vol. 1, no. 3, pp. 213–219, Aug. 1990, doi: 10.1287/orsc.1.3.213.

[94]. S. Mendonça, T. S. Pereira, and M. M. Godinho, "Trademarks as an indicator of innovation and industrial change," Research Policy, vol. 33, no. 9, pp. 1385–1404, Nov. 2004, doi: 10.1016/j.respol.2004.09.005.

[95]. L. Argote and P. Ingram, "Knowledge Transfer: A Basis for Competitive Advantage in Firms," Organizational Behavior and Human Decision Processes, vol. 82, no. 1, pp. 150–169, May 2000, doi: 10.1006/obhd.2000.2893.

[96]. R. S. Sikes, "2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education:," Journal of Mammalogy, vol. 97, no. 3, pp. 663–688, May 2016, doi: 10.1093/jmammal/gyw078.

[97]. S. Offsey, "Knowledge Management: Linking People to Knowledge for Bottom Line Results," Journal of Knowledge Management, vol. 1, no. 2, pp. 113–122, Jun. 1997, doi: 10.1108/eum0000000004586.

[98]. A. S. Bollinger and R. D. Smith, "Managing organizational knowledge as a strategic asset," Journal of Knowledge Management, vol. 5, no. 1, pp. 8–18, Mar. 2001, doi: 10.1108/13673270110384365.

[99]. P. Rikhardsson and O. Yigitbasioglu, "Business intelligence & analytics in management accounting research: Status and future focus," International Journal of Accounting Information Systems, vol. 29, pp. 37–58, Jun. 2018, doi: 10.1016/j.accinf.2018.03.001.

[100]. H. Hassani, C. Beneki, S. Unger, M. T. Mazinani, and M. R. Yeganegi, "Text Mining in Big Data Analytics," Big Data and Cognitive Computing, vol. 4, no. 1, p. 1, Jan. 2020, doi: 10.3390/bdcc4010001.

[101]. D. Yan, K. Li, S. Gu, and L. Yang, "Network-Based Bag-of-Words Model for Text Classification," IEEE Access, vol. 8, pp. 82641–82652, 2020, doi: 10.1109/access.2020.2991074.

[102]. W. A. Qader, M. M. Ameen, and B. I. Ahmed, "An Overview of Bag of Words;Importance, Implementation, Applications, and Challenges," 2019 International Engineering Conference (IEC), pp. 200–204, Jun. 2019, doi: 10.1109/iec47844.2019.8950616.

[103]. S.-Y. Yang, "An ontological website models-supported search agent for web services," Expert Systems with Applications, vol. 35, no. 4, pp. 2056–2073, Nov. 2008, doi: 10.1016/j.eswa.2007.09.024.

[104]. R. Graham, "Jewish community education: continuity and renewal initiatives in British Jewry 1991-2000," 2011. [Online]. Available: http://eprints.hud.ac.uk/id/eprint/14070/

[105]. P. Zschech, R. Horn, D. Höschele, C. Janiesch, and K. Heinrich, "Intelligent User Assistance for Automated Data Mining Method Selection," Business & Information Systems Engineering, vol. 62, no. 3, pp. 227–247, Mar. 2020, doi: 10.1007/s12599-020-00642-3.