A Semiautonomous Approach for Strategic Decision-Making; an Application for Digital Transformation in Finance

Peiman Alipour Sarvari (peiman.alipour@list.lu)
Luxembourg Institute of Science and Technology
Sebastien Martin
Royal Melbourne Institute of Technology
Andrius Grybauskas
Kaunas University of Technology
Gulcan Baskurt
Beykent University

Research Article

Keywords: Technology roadmapping, Finance, artificial intelligence, parallel coordinates, neutrosophic fuzzy sets, particle characterise analysis

Posted Date: September 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2008989/v1

Title: A Semiautonomous Approach for Strategic Decision-Making; an Application for Digital Transformation in Finance

Peiman Alipour Sarvari (Corresponding author)
1. Luxembourg Institute of Science and Technology, Luxembourg
2. School of Business and Economy at Kaunas University of Technology, Lithuania
e-mail: peiman.alipour@list.lu

Sebastien Martin
Royal Melbourne Institute of Technology, Australia
e-mail: sebastien.martin@rmit.edu.au

Andrius Grybauskas
School of Business and Economy at Kaunas University of Technology, Lithuania
e-mail:andrius.grybauskas@ktu.lt

Gulcan Baskurt
School of Business Management, Beykent University, Turkey
e-mail: glcn_20@hotmail.com

Abstract:
This paper proposes an innovative approach to help domain knowledge experts with technology roadmapping and the strategic decision-making process. In our approach, we introduce a retrofitted parallel coordinate visualization tool, which is enhanced with a feature importance analysis algorithm. Our approach has produced four significant achievements. First, our new method reveals the shortcomings of some traditional parallel coordinate tools. Second, our method applies neutrosophic fuzzy sets in tacit knowledge analysis to foster ease with written text data; it assists with the processes in exploring possible roadmapping strategies, measuring vicinity conditions of adjacent features, and prioritizing their order based on the desired solution space of a certain attributes. The methods are applied in a case study with actual data that comes from a research project designed for financial organizations' digitalization. Third, the application of parallel coordinates is reported for the first time for technology roadmapping in finance. Finally, proof of concept and a benchmarking practice considering generated hypotheses are provided in this paper.

Keywords: Technology roadmapping, Finance, artificial intelligence, parallel coordinates, neutrosophic fuzzy sets, particle characterise analysis

Statements and Declarations

Conflict of interest

The authors declare that they have no conflict of interest regarding the publication of this article.

Ethical approval

This article does not contain any sensitive data, personal data, studies with human participants or animals performed by any of the authors.

Acknowledgment
This research is financed by the European Union's Horizon 2020 research and innovation programme under grant agreement No 810318 (Project "Industry 4.0 impact on management practices and economics (IN4ACT)".

Title: A Semiautonomous Approach for Strategic Decision-Making; an Application for Digital Transformation in Finance.

Abstract:
This paper proposes an innovative approach to help domain knowledge experts with technology roadmapping and the strategic decision-making process. In our approach, we introduce a retrofitted parallel coordinates visualization tool, which is enhanced with a feature importance analysis algorithm. Our approach has produced four significant achievements: First, our new method reveals the shortcomings of some traditional parallel coordinate tools. Second, our method applies neutrosophic fuzzy sets in tacit knowledge analysis to foster ease with written text data; it assists with the processes in exploring possible roadmapping strategies, measuring vicinity conditions of adjacent features, and prioritizing their order based on the desired solution space of a certain attributes. The aforementioned methods are applied in a case study with actual data that comes from a research project designed for financial organizations' digitalization. Third, the application of parallel coordinates is reported for the first time for technology roadmapping in finance. Finally, proof of concept and a benchmarking practice considering generated hypotheses are provided in this paper.

Keywords: Technology roadmapping, Finance, artificial intelligence, parallel coordinates, neutrosophic fuzzy sets, particle characterise analysis

1 Introduction

With the increasing volume and variety of information available, decision-makers need to effectively manage large, multivariate data. Thus, decision-making in management has become a matter of primary importance among today's business issues. Indeed, some scholars have gone so far as to equate management with decision-making [1], because the manager must make decisions continuously while performing the functions of planning, organizing, directing, coordinating and controlling the management works [2]. In other words, decision making is an element that forms the basis of all management activities rather than being a separate management activity [3]. Making decisions requires careful thought; however, there is more than one type of thinking. Where there is only one "correct" solution, an elimination-oriented thinking style is needed [4]; In situations where new ideas are expected to be produced, a creative way of thinking that brings ideas other than known solutions to the agenda should be adopted [5]. Strategic Decision-Making (SDM) is considered with the potential to not only maintain the current position but also create the future position in the competition [6]. However, creating the future position involves a difficult process, or SDM process that is not easy to express in terms of theory and practice [7]. The main reason for this is the multiplicity, variety, and quality of the factors affecting the process. Increasing change with the developing technology, intense competition, and increasing uncertainty forces new proposals in decision making. SDM emerges as one of these propositions [8]. Strategic decisions are long-term future-oriented and result-oriented decisions where change and competition are intense, and risk, complexity, and uncertainty are high. Furthermore, when said strategic decisions are made in order to obtain a more advantageous situation/position than competitor activities [9]. Simply put, SDM is explained as the decisions made for the far/long future where risk, complexity and uncertainty are acute. This naturally requires strategic decisions to be creative, innovative, change-oriented, sustainable, inclusive, diverse and effective.

Faced with digital world reality, executives, senior executives, and CEOs admit that digital technology can significantly improve the performance of existing businesses [10]. Furthermore, many have started the transformations by changing the culture and processes of the organization, pondering which digital tools to use and how to use them [11]. Those who took the first step in digital technologies to make their old business models were more innovative and were more advantageous to meet their customers' needs [12].

Programmed decisions are repetitive and routine, with established rules and procedures (often automated), involving objects rather than people. They can be migrated to lower levels of the organization. For example, stock control, vehicle downloads and listing are classic examples of programmed decisions. In contrast, unprogrammed decisions, are innovative and non-routine. They have no decision rules, contain a high degree of uncertainty, cannot be pushed down to lower levels, and are related to objects—but predominantly involve people [13]. For example, acquisitions, mergers, launching a new product, and staffing of unprogrammed decisions. There are some relationships between management levels and decision types. Although there is no absolute rule, programmed decisions are made at lower levels and less programmed decisions are made at higher levels of management [14].

Improving the available background towards the desired foreground with a targeted Technology Readiness Level (TRL) requires a clear understanding and estimation of available resources for example: 1) dealing with security concerns in both physical and cyber; needed further enhancement for Systems integration; 2) improved managing multi-vendor solutions; ensured network stability/reliability; dealing with skills shortage; tailored training requirements; precise size estimations of financial investment or uncertain ROI; 3) dealing with privacy concerns (GDPR); 4) reducing fear among users to adopt new technologies (developing a cloud strategy, aligning telecom with business strategies, ensuring regulatory compliance, tactical solution to overcome overwhelming tasks due to volume of data), and 5) addressing the lack of standards and interoperability with defining scopes of TRL and life span. The mentioned drivers are evaluated based on the following key-questions;

- We are on track with others in our industry
- We are pioneering others in our industry
- We are behind others in our industry

Past research on developing and changing technologies, market direction of products or services, and the choice of innovative business methods has tried to help organizations plan their technological development in different areas and provide a tool for accurate technology roadmap design. Many of the previous road mapping processes have focused on improvement and optimization.

Sarvari et.al. in their work provides domain literature through the review of the stages of digital transformation as identified by academic literature [15]. Regardless of auditing tools, the scope of the audit must be in accordance with the market position and the activities of the company. As a result of the trends and changes that have emerged today, it has become very difficult to rely on the trial-and-error approach while performing the management action, especially when making decisions. Managers must now become more sophisticated and learn how to use new tools and techniques developed in their area of expertise. Decision-making is essentially a problem-solving tool. In this respect, it is necessary to clearly know what the problem is, its importance, scope, impact and consequences. For example, illness is a health problem. To cure the disease, it is imperative to make a clear and precise diagnosis of it. Sometimes, it can be concluded that a condition that is thought to be a disease is not so important that it requires attention, at least for the time being. Precautionary measures against an insignificant event may cause an undesirable situation with unpredictable results. A basic and objective decision-making process begins with the definition of the problem. For an accurate definition of any problem, we must first gather information about the background of the issue. Depending on the problem, information can be gathered by searching relevant literature, interviewing affected personnel, consulting external experts, or examining past records. Therefore, correct diagnosis of the situation is imperative for sound decision making [17].

In this regard, the importance of written resources and text mining is more and more revealed recently. Knowledge discovery, information extraction and statistical pattern learning are some of the practices that scholars and data scientists applying on the text data [17]. Natural Language processing is another deep learning technique that tries to understand the context and meaning of a text emulating human

[18]. However, transforming text to numerical matrixes is not enough to understand the true meaning of the written text. The reality of tacit knowledge is the main impediment on the way of interpreting the true meaning of a context so good that one can consider it as explicit information [19].While there is a large body of knowledge concerning the broad framework or focuses on specific tasks of digital transformation, however the processes that a company might use to stay aware of the components of currently available and potential disruptive technologies are slow [20]. In this perspective, an information watch led internally or with the contribution of consultants, is a fundamental process to let the company know (and then score its achievements) relatively the state-of-the-art technologies and its competitors. For this, a digital radar plot, showcased in [21] is applied to represent the relevant useful landscape of technologies available at a given time.

In addition to the challenges caused by tacticity of context, the decision-making process need to overcome complexity of data due to their multidimensionality that is known as multivariate data. This difficulty of interpretation and visualization of multivariate data is getting worse with the growth of data [22]. Hence, the use of parallel coordinates (PCs) plots is extremely popular among data analysis experts. Thanks to PCs plots, one can visualize and interpret the relationships among multiple feature and factors those are shaping the datasets simultaneously, regardless of their size or number of attributes. While it is just possible up to three attributes in a Cartesian system [23]. Nevertheless, PCs plots have their specific problems and use-difficulties like clutter of the plot when the numbers of feature increases, that make interpretation of the plot almost impossible. Another issue with them is that the order of neighbor features on the plot is predefined and one need to move the axes to find the best correlation among features [24]. Despite the popularity of their applications, PCs plots are mainly unknown among managers and decision-making experts. One can read more information about pros and cons of using PCs plots on the work of Alminagorta et al [25].

In this study, we present a semi-autonomous innovative, yet methodological approach, to use the most of both consensus and literature data. Acknowledging the fact, the domain expert knowledge and domain context from the literature could shape a prosperous repository of factors to shape decision making hypotheses and technology roadmapping scenarios by the final decision makers. However, the raw text does not always necessarily provide a lean and explicit knowledge and mainly, large number of statements, terms and contexts holding knowledge as fuzzy and tacit. Therefore, a selection of related terms out of the abundance of text data is the matter of tacit knowledge extraction. In addition to that, for the first time in the literature, we will present ~~an~~ innovative PCs tool to explore multidimensional yet subjective decision data. The proposed PC tool is equipped with an evolved principal component analysis (PCA) algorithm that reduces the clutter of the PCs to facilitate the corresponding interpretations. Nevertheless, this hybridization, makes help to autonomously find the best order of adjacent and nonadjacent features in a typical PCs graph which currently is considered as one of the main shortcomings of traditional PCs methods. The core concepts and key features of the proposed methodology is going to be applied in a strategic decision-making problem in finance industry that is handled for the first time in the literature by this paper. A comprehensive research survey fulfilled by domain knowledge experts (DKE) is applied to provide deep understanding of decision factors to solve “digitalization roadmapping in finance” problem.

The organization of this paper starts with a comprehensive **introduction** emphasizing the importance of decision making and the need to contribute to the development of assistive technologies and adaptive tools to help decision makers. Then, the proposed methodology at **section three** follows the introduction over **methods** and techniques at **section two**, where the developed PCs and tacit knowledge extraction have been introduced. Also, in this section we have reviewed backgrounds of each method and technology in their related literature. At **section four**, we start to showcase the execution of the methodology and our new PCs tool with a benchmarking attempt. At **section five**, the paper deeply discusses the results using logical interpretation from managerial and digitalization perspective in finance industry with the help of DKEs.

2 Methods

2.1. Background of text mining

Text is one of the most common sources of data available ubiquitously. Therefore, benefitting from such a rich source of data requires tools and algorithms like text data mining that is mainly known as text mining [26]. Text mining is a collection of techniques and algorithms to conduct a process to transform unstructured format of data into structured data to extract meaningful insights out of them [27]. According to literature, nowadays, large-scale literature reviews are being conducted using text mining approaches in the different sciences and sources of data. Natural Language Processing (NLP) and frequency analysis can reveal emerging, sparsely addressed areas of interest in managerial sciences. In this way, scientists and data engineers apply classification techniques such as Support Vector Machines (SVM), Naïve Bayes and other advanced deep learning and analytical techniques to recognize patterns among unstructured textual data. The target is to deliver numerical matrixes at the end [28].

According to Hartmann et al., 2019, text mining is divided into two classes of machine learning (ML) and lexicon-based methods [29]. In order to make the best at selecting either of these approaches, one should be equipped with the subjective knowledge over research objectives as well as computational and data analysis skills. To succeed using ML techniques in text mining processes, one needs to have larger portion of numerical skills over subject matter skills. ML based text mining helps with benefit of tremendous scalability for large-size projects and research. Whereas, lexicon-based methods, are better fit to small-scale companies and projects as they require establishing dictionaries of terms and more intuitive means to analyse the text [30]. There are plenty of text mining approaches available at the works of Michael et al., [31] and Feng et al., [32] and quite comprehensive number of techniques, methods, tools, libraries and packges that one can get more information reading the work of Ignatow and Mihalcea [33]. The neutrosophic term weighing technique developed by Bounabi et al., is considered as the main text mining analysis method used to execute to be defined methodological approach [34].

2.2. Background of tacit knowledge applications

Contrary to explicit knowledge, Tacit knowledge refers to the type of knowledge that cannot be codded or expressed clearly in language. If there is an interaction among technology, people and processes, there is a knowledge management application. According to Brooking, [35], knowledge management is the approach to better sharing, managing, and organizing communication assets those have human in the centre.

Polanyi believes, people know more than what they can tell. He gives the example of recognizing a face among millions of faces by a person, that is difficult to put the reason, the process and the logic of such recognition on the paper, but what is an obvious fact is that the person has recognized the “face as a whole” [36]. To support this statement, Alavi and Leidner described the difference between tacit and explicit on their research [37]. Briefly, as converse of explicit knowledge, tacit knowledge is mainly defined by Nonaka and Takeuchi as personal believes, knowledge, individual thoughts, experiences and perspectives about and around that person’s ecosystem. Unlike explicit knowledge, tacit knowledge could be subdivided, re-formulized, and subjected to multiple interpretations [38]. To support this idea, the example from Gordon’s book about the Bessemer steel process would show the difficulties of describing rules on the paper let alone, describing sensitive justifications in shape of words. Henry Beessmer sold a patent that was describing a process of an advanced annealing. However, the purchasers couldn’t get this patent to work despite its compliance with patent writing standards. After a while, Henry Beessmer, established his own steel annealing company and succeeded applying the exat contents of the patent. This example showcases that narrow difference among being informed and having knowledge [39]. One can read more about this topic on Davis et., (2015) paper [40]. In the current paper, we will use a single valued neutrosophic sets (SVNS) technic to extract tacit knowledge from the available text.

2.3. Background of PCs

Exploring multivariate data is a challenging task where Cartesian systems are unable to adequately demonstrate the quantification and identification among interrelated variables with more than 3 dimensions, unless one uses multivariate analyses techniques those are statistics-based covariation summarizers [41]. To overcome this difficulty, the data could be represented as a polyline and where each dimension and features of data could be associated with a vertical axis. This way, each vector or element could be projected on a graphical way pointing out across the vertical indicators. Furthermore, it is very straightforward to recognize positive correlations (parallel lines) and negative correlations (crossed intersected in a point) in a mapped PCs plot [42]. As it is demonstrated on Fig. 1, PCs could represent Cartesian coordinates ellipses as hyperbolas [43]. Obviously, Cartesian systems are limited with the numbers of axes where, the advantage of PCs is that they can condense a large quantity of data and present them with numerous axes in an interactive manner. Furthermore, various features of plot interactivity can be applied to facilitate the identification of patterns and the formation of hypotheses concerning relationships among multivariate data [44].

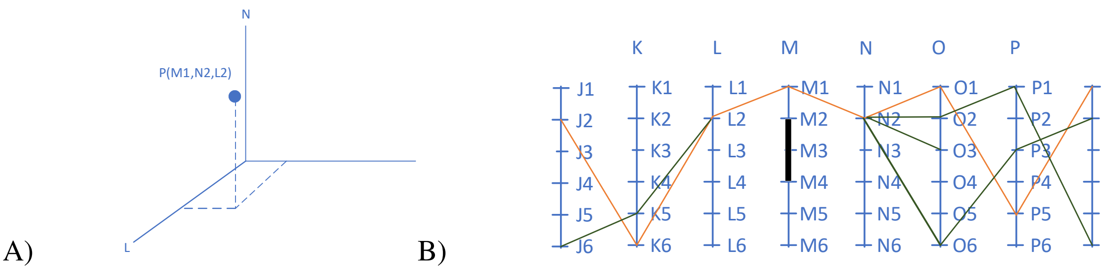


Fig. 1. (A) Cartesian system coordinating a point (B) A point is coordinated as a line in PCs plot

According to the literature, one of the main shortcomings of PCs technique is that when the size of data grows, it gets too cluttered and it reduces the possibility of sound interpretation. Obviously the clutter of the connected coordinates is reducing the applicability of the tool. Thus, the interconnectivity of one feature (M) is naturally but inadequately getting observed just with their adjunct neighbour features (L, N). That is the main reason that in many PCs tools and packages, they ask the user to play with the order of the adjacent features to find the most meaningful visualization and feature relationship. While the current PCs techniques have difficulty to illustrate the proper interconnectivity of a feature with their nonadjacent features (J,K,P,Q, etc), let alone, to show the interconnectivity of features while we target a certain solution space of a feature (i.e., M2-M3). As a solution to this shortcoming of PCs, we propose the use of a state-of-the-art dynamic PCA embedded under the neath of a PCs visualization tool.

The principal of PCA algorithms is based on the fact that each component of data is a projection of one or more than one feature of a dataset. Reducing the number of features would be acceptable unless, it negatively impacts the representativeness of the solution space's largest eigenvalues. However, the ability of PCA is not necessarily limited to feature importance analysis. It can be used to find the most valuable features or portion of features those are impacting a range of solutions on another feature. This characteristic could tackle the above-mentioned shortcomings of PCs. Hereunder, we introduce a dynamic PCA algorithm that can only perform under a PCs tool environment. The main feature of the proposed PCs is its unique colour spectrum-based visualization that is using a weighing approach. We will call it "comprehensive factors" selection algorithm that is using the eigenvalues extracted from PCA to assign colour to each connection among features.

Let's consider the followings as the components of our analysis where x and y representing a range on a feature between minimum and maximum amounts of that feature.

Data matrix: $K_{m\times n}; m = \{m_1, m_2, \dots, m_{1m}\}, n = \{n_1, n_2, \dots, n_n\}$

Feature's empirical mean set: $\mu = \{\mu_{n1}, \mu_{n2}, \dots, \mu_{nn}\}$

Solution space on each feature: $R_n^{x,y};\ n = \{n_1, n_2, \dots, n_n\},\ x = [n_{min}, n_{max},],\ y = [n_{min}, n_{max}]$

Coefficients vector: $W_i = \{w_1^{x,y}, w_2^{x,y}, .., w_n^{x,y}\};\ n = \{n_1, n_2, \dots, n_n\},\ x = [n_{min}, n_{max},],\ y = [n_{min}, n_{max}]\ ;\ i = \{1, \dots, n\}$

Principal component scores: $S_i = \{s_1^{x,y}, s, .., s_n^{x,y}\};\ n = \{n_1, n_2, \dots, n_n\},\ x = [n_{min}, n_{max},],\ y = [n_{min}, n_{max}];\ i = \{1, \dots, n\}$

$$w_i^{x,y} = \arg\max\left\{\sum_{n=1}^{n} \int_x^y s_i^{x,y} \times |\mu_n|\right\}$$

$$\mu_i = \mu - \sum_{n=1}^{k-1} \mu_n \times (w_n^{x,y})^T$$

$$S = w_i^{x,y} \times \mu_i$$

After calculating each features principal component score against other features as well as different ranges of solution space on a certain feature, it is turn of extracting the covariances among bilateral j and k features impacted by their W_i eigenvalue properties.

Covariance: $cov_{j,k} = \sum_i (\mu_i \times \arg\max\{\sum_{n=1}^{n} \int_x^y s_i^{x,y} \times |\mu_n|\})^2 \times (\arg\max\{\sum_{n=1}^{n} \int_x^y s_j^{x,y} \times |\mu_n|\})^T \times \arg\max\{\sum_{n=1}^{n} \int_x^y s_k^{x,y} \times |\mu_n|\}$

The next step after calculating the covariance is Eigen Value Decomposition of the Covariance matrix, that is $S_M = \mu_n \times w_M^{x,y}$ as the matrix S_M now has n rows and just M columns. The dimensionality reduction based on the ordering of the bilaterally calculated covariances happens regarding the eigenvalues listed in the S_M.

3 Methodology

In this section we introduce a methodology to execute an automated and semi-autonomous technology roadmapping tool to assist strategic decision makers. No doubt that there are experts in companies and firms to hep strategic decision makers on the way of digitalization decisions such as technology selections, target projection and process definition. Nevertheless, there are plethora of written papers, articles, books, and reports those holding valuable explicit and tacit knowledge.

On the other hand, with the emergence of digital transformations, Fintech initiatives have started to spread to a wider area. FinTech organizations and start-ups have brought a new perspective to the financial services industry with their agility, flexibility, forward-looking strategies and new customer-oriented business models. It reshaped the finance and banking sector as the number of Fintech enterprises, especially in the USA and Europe, gained more momentum as of the 2008 financial crisis [45]. Based on this and regarding demonstrate the applicability and functionality of text mining, tacit knowledge extraction and PCs to financial enterprises and related research, we introduce a **case study** with a unique derivate from a digitalization project. Fig. 2 depicts the used methodology in this paper. Due to the rapid development and transformation of financial services, this article proposes a structured and dynamic method of hypothesis creation, decision making and technology roadmapping in financial organizations that is consisting of three phases.

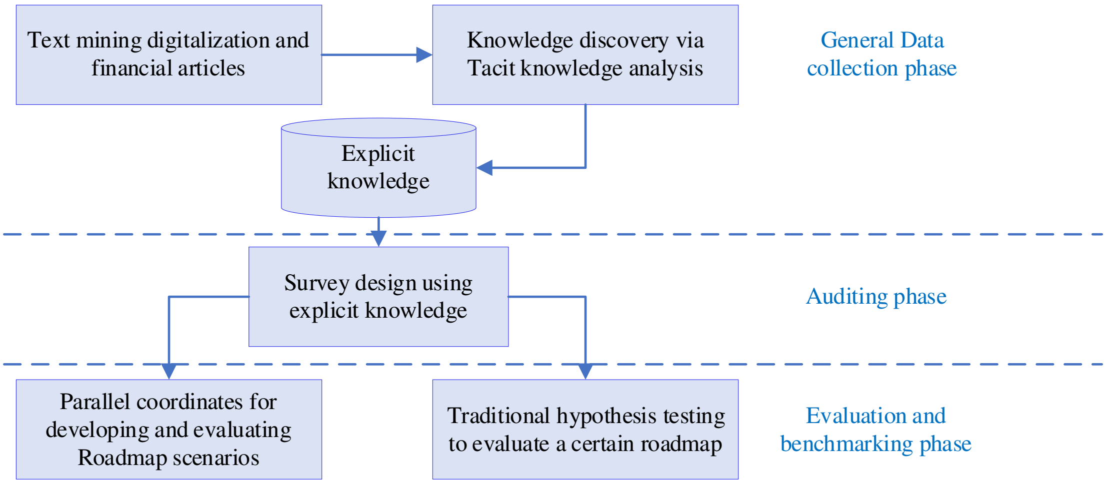


Fig. 2. Research methodology and proposed approach

Phase 1; General data collection phase: it conducts the value definition; requirement collection and preparation need basing on the available literature and technologies. This phase will use a global hybrid template of regional and global requirements to achieve defined goals and ambitious targets via an investigation strategy. It is using data collection, text mining and tacit knowledge extraction techniques to deliver an explicit knowledge dataset to be used in the next phases. In this phase the first task is applying a web-scrapping algorithm using all major keywords related to "digitalization", "finance" and lexicon repository provided by University of California's "Glossary of Digital Library Terms"[1] using "Python programming"[2] language and "Knime software"[3] those are open-source tools for programming. The scrapping is applied on financial news of over 30 online newspapers as well as more than 582 documents downloaded from "ScienceDirect"[4] database. The downloaded documents are articles, books, reports published between 2018-2023.

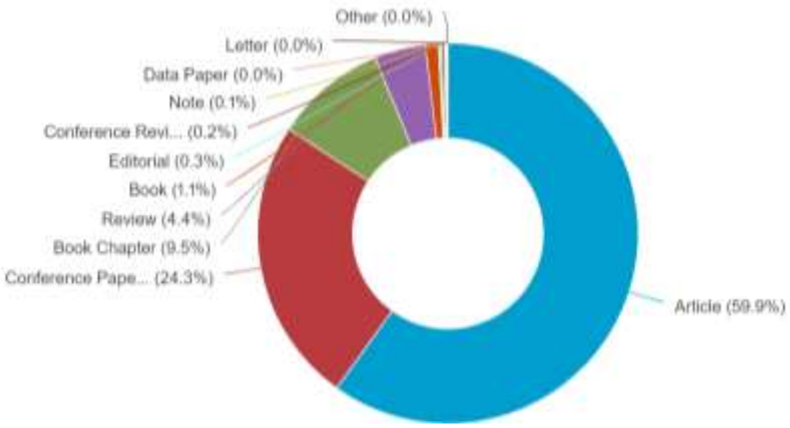
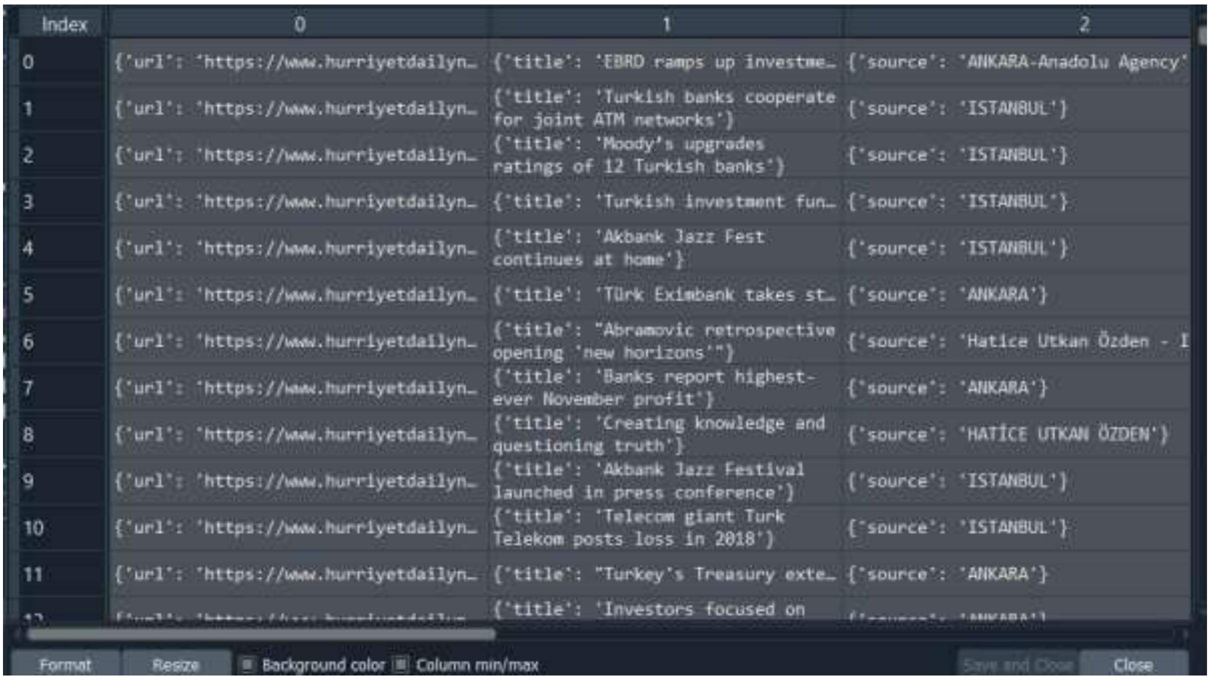


Fig. 3. Downloaded research documents by type from "ScienceDirect"

Fig. 3, showing the types of the documents downloaded from ScienceDirect while Fig. 4 is a screenshot partially showing the sources of the web scrapped newspapers. The "Scrapy"[5] and "Selenium"[6] python packages those are open-source and collaborative frameworks are used to extract data by web scrapping.

---

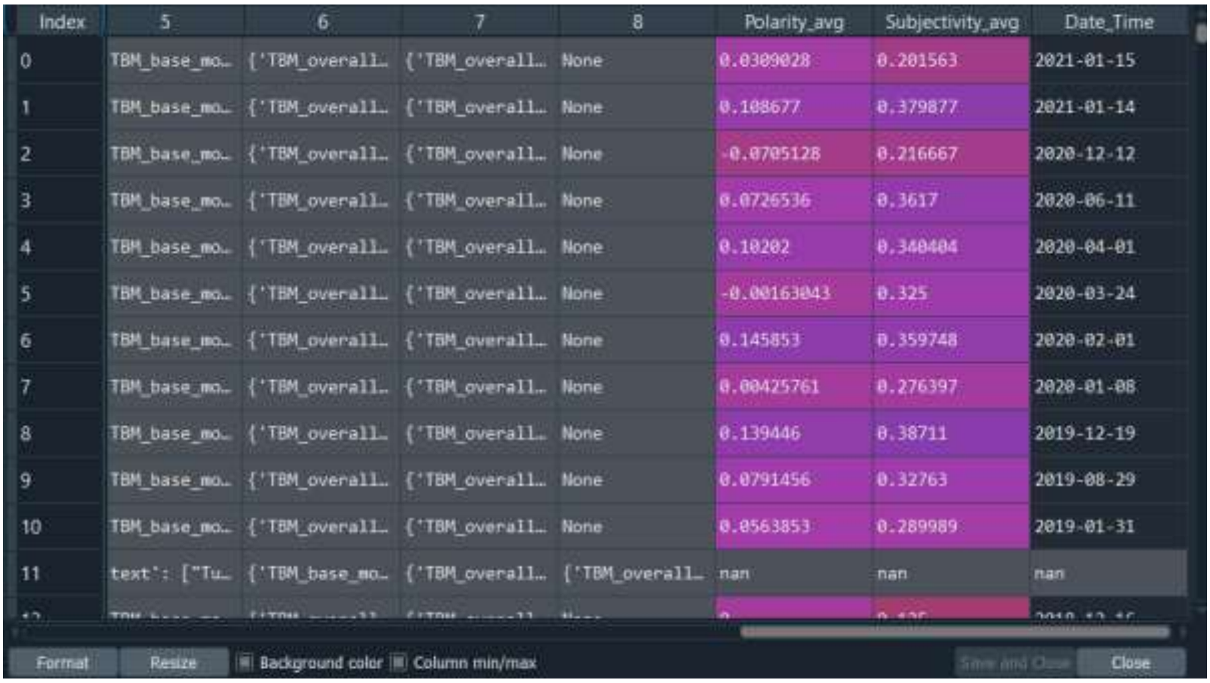
[1] https://cdlib.org/resources/technologists/glossary-of-digital-library-terms/
[2] https://www.python.org/
[3] https://www.knime.com/
[4] https://www.sciencedirect.com/
[5] https://scrapy.org/
[6] https://www.selenium.dev/



Index	0	1	2
0	{'url': 'https://www.hurriyetdailyn…	{'title': 'EBRD ramps up investme…	{'source': 'ANKARA-Anadolu Agency'
1	{'url': 'https://www.hurriyetdailyn…	{'title': 'Turkish banks cooperate for joint ATM networks'}	{'source': 'ISTANBUL'}
2	{'url': 'https://www.hurriyetdailyn…	{'title': "Moody's upgrades ratings of 12 Turkish banks"}	{'source': 'ISTANBUL'}
3	{'url': 'https://www.hurriyetdailyn…	{'title': 'Turkish investment fun…	{'source': 'ISTANBUL'}
4	{'url': 'https://www.hurriyetdailyn…	{'title': 'Akbank Jazz Fest continues at home'}	{'source': 'ISTANBUL'}
5	{'url': 'https://www.hurriyetdailyn…	{'title': 'Türk Eximbank takes st…	{'source': 'ANKARA'}
6	{'url': 'https://www.hurriyetdailyn…	{'title': "Abramovic retrospective opening 'new horizons'"}	{'source': 'Hatice Utkan Özden - I
7	{'url': 'https://www.hurriyetdailyn…	{'title': 'Banks report highest-ever November profit'}	{'source': 'ANKARA'}
8	{'url': 'https://www.hurriyetdailyn…	{'title': 'Creating knowledge and questioning truth'}	{'source': 'HATİCE UTKAN ÖZDEN'}
9	{'url': 'https://www.hurriyetdailyn…	{'title': 'Akbank Jazz Festival launched in press conference'}	{'source': 'ISTANBUL'}
10	{'url': 'https://www.hurriyetdailyn…	{'title': 'Telecom giant Turk Telekom posts loss in 2018'}	{'source': 'ISTANBUL'}
11	{'url': 'https://www.hurriyetdailyn…	{'title': "Turkey's Treasury exte…	{'source': 'ANKARA'}
12	[illegible]	{'title': 'Investors focused on	[illegible]

Fig. 4. A screenshot showing partially the sources of the newspapers and journals

To mine the text, we needed to explore the Polarity and Subjectivity of each statement mentioned on the downloaded text files. We used the methodology described on the work of Piao et al., [46]. The corresponding results for Polarity and Subjectivity[7] are duly demonstrated on Fig. 5 that is a screenshot depicting polarity and subjectivity measures of each statement on downloaded files.



Index	5	6	7	8	Polarity_avg	Subjectivity_avg	Date_Time
0	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.0309028	0.201563	2021-01-15
1	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.108677	0.379877	2021-01-14
2	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	-0.0705128	0.216667	2020-12-12
3	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.0726536	0.3617	2020-06-11
4	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.10202	0.340404	2020-04-01
5	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	-0.00163043	0.325	2020-03-24
6	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.145853	0.359748	2020-02-01
7	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.00425761	0.276397	2020-01-08
8	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.139446	0.38711	2019-12-19
9	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.0791456	0.32763	2019-08-29
10	TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	None	0.0563853	0.289989	2019-01-31
11	text': ["Tu…	{'TBM_base_mo…	{'TBM_overall…	{'TBM_overall…	nan	nan	nan

Fig. 5. A screenshot depicting polarity, subjectivity and recency measures of each statement on each file

[7] Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1].

The second task of the first phase is applying "neutrosophic term frequency – inverse term frequency (NTF-IDF)", that "is an extended version of the popular fuzzy TF-IDF (FTF-IDF) and uses the neutrosophic reasoning to analyze and generate weights for terms in natural languages" as a text mining approach [47]. We use the hereunder formulation described on the work of Majumdar to extract explicitly out of tacticity using, the extracted polarity, subjectivity, and recency matrix. The algorithm is applied to select the best fitting terms and contexts to the " decision making processes and digitalization factors related to finance", for the first time in this paper

Let X be a universal set. A Neutrosophic set A in X is characterized by a truth-membership function t_A, a indeterminacy-membership function i_A and a falsity-membership function f_A, $where\ t_A, i_A, f_A: X \to [0,1]$, are functions and $\forall x \in X$, $x \equiv x(t_A(x), i_A(x), f_A(x)) \in A$, is a single valued neutrosophic element of A [48].

A single valued neutrosophic set A (single valued neutrosophic sets in short) over a finite universe $X = \{x_1, x_2, x_3, \dots, x_n\}$ is represented as

$$A = \sum_{i=1}^{n} \frac{x_i}{< t_A(x_i), i_A(x_i), f_A(x_i) >}$$

Where; $t_A(x_i)$ is Polarity, $i_A(x_i)$ is Subjectivity and $f_A(x_i)$ is Recency.

Now that the point A is defined, lets elaborate it in terms of "Truth" and "Falsity" components.

The complement of a single valued neutrosophic set A is denoted by A^c and is defined by

$$t_{A^c}(x) = f_A(x);\ i_{A^c}(x) = 1 - i_A(x) \&\ f_{A^c}(x) = t_A(x)\ \forall x \in X.$$

A single valued neutrosophic sets A is contained in the other single valued neutrosophic set B, denoted as $A \subset B$, if and only if $t_A(x) \le t_B(x); i_A(x) \le i_B(x) \& f_A(x) \ge f_B(x)\ \forall x \in X$. The two sets will be equal, i.e. $A = B$, if $A \subset B \& B \subset A$.

Let us denote the collection of all single valued neutrosophic sets in X as $N(X)$.

Several operations like union and intersection have been defined on a single valued neutrosophic sets and they satisfy most of the common algebraic properties of ordinary sets. On the other hand, C=AUB and is defined as:

$$t_C(x) = max(t_A(x), t_B(x)); i_C(x) = max(i_A(x), i_B(x)) \& f_C(x) = min(f_A(x), f_B(x))\ \forall x \in X.$$

The $A \cap B$ is a single valued neutrosophic set C:

$$t_C(x) = min(t_A(x), t_B(x)); i_C(x) = min(i_A(x), i_B(x)) \& f_C(x) = max(f_A(x), f_B(x))\ \forall x \in X.$$

Let's define 'truth favourite' and 'falsity favourite' operators those will help with removing indeterminacy in the single valued neutrosophic set by transforming it to an Intuitionistic Fuzzy Set.

The truth favourite of a single valued neutrosophic sets is again a single valued neutrosophic sets B written as $B_\Delta A$ which is defined as follows:

$$T_B(x) = min(T_A(x) + I_A(x), 1)$$
$$I_B(x) = 0$$
$$F_B(x) = F_A(x),\ \forall x \in X.$$

The falsity favourite of a single valued neutrosophic sets A is again a single valued neutrosophic sets B written as $B = \nabla A$, which is defined as follows:

$$T_B(x) = T_A(x)$$
$$I_B(x) = 0$$
$$F_B(x) = min(F_A(x) + I_A(x), 1),\ \forall x \in X.$$

If A, B, C are three single-valued neutrosophic sets then the following holds:

$$(i) A \cup B = B \cup A; A \cap B = B \cap A$$
$$(ii) A \cup (B \cup C) = (A \cup B) \cup C;\ A \cap (B \cap C) = (A \cap B) \cap C$$
$$(iii) A \cup (B \cap C) = (A \cup B) \cap (A \cup C);\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$(iv) A \cup A = A;\ A \cap A = A;\ \Delta\Delta A = A; \nabla\nabla A = A$$
$$(v) A \cup (A \cap B) = A;\ A \cap (A \cup B) = A$$
$$(vi) (A \cup B)^C = A^C \cap B^C; (A \cap B)^C = A^C \cup B^C.$$

The last task of phase 1, is to collect the explicit part of the knowledge while dealing with probable implicitness. To handle this task the knowledge diffusion methodology presented above has been applied on the data and the first 11 rows of the results is presented on Table 1. According to the total score statement 1, 6 and 8 have been selected as the "knowledge related to digitalization" referencing "Glossary of Digital Library Terms" lexicon repositories.

	Polarity	Subjectivity	Recency	A	B	C	Total Score
0	0,03090	0,20156	44211	0,001408889	0,498591111	0,298591111	0,601695445
1	0,10868	0,37988	44210	0,009338098	0,490661902	0,290661902	**0,611418981**
2	-0,07051	0,21667	44177	-0,00345831	0,50345831	0,30345831	0,595878534
3	0,07265	0,36170	43993	0,005973414	0,494026586	0,294026586	0,607254768
4	0,10202	0,34040	43922	0,007906764	0,492093236	0,292093236	0,609640568
5	-0,00163	0,32500	43914	-0,000120666	0,500120666	0,300120666	0,599855236
6	0,14585	0,35975	43862	0,011962583	0,488037417	0,288037417	**0,614706967**
7	0,00426	0,27640	43838	0,000268441	0,499731559	0,299731559	0,600322302
8	0,13945	0,38711	43818	0,012319376	0,487680624	0,287680624	**0,615156694**
9	0,07915	0,32763	43706	0,005932942	0,494067058	0,294067058	0,607205025
10	0,05639	0,28999	43496	0,003759224	0,496240776	0,296240776	0,604545242

Table 1. Scoring Truth Favorite of each statement appeared on Fig. 5.

Phase 2; Auditing phase: the target of this phase is to successfully implement a survey to collect the views of DKE. It shapes the survey of targeted collocutors, and questionnaires to evaluate the probable foreground, available background and the technologies in the enterprise via technical requirements analyses questions embedded in the questionnaires.

The first task of this phase is to shape a focus group encompassing people with proper level of knowledge about the organization as well as the trended background in the literature and other competitor organizations. In terms of testbed for this research, we used a survey fulfilled by 2853 experts in multiple financial enterprises. The survey is composed from numerous backgrounds, foregrounds, technologies and needed processes to realize a strategic scenario plan. Each focus group is supposed to select the most important case from their own point of views, and each composed of 400 experts from the experts' list illustrated on Table 2. Dividing the experts to multiple focus groups will make help to define common ground truth and sanity check on survey results and hypotheses tests.

Job title	**Numbers**	**Job title**	**Numbers**	**Job title**	**Job title**
Data scientist	172	Integration& support manager	89	Cybersecurity technician supervisor	61
Senior specialist	80	Diffusion technique and methodological development expert	25	Director of systems and operations	14
Digitalization team leader	59	Computer systems division expert	187	Data janitor	103

It official	109	Financial management technician	124	Bi analyst	153
Expert associate for controlling	95	Computer systems technician	52	Customer relationship associate	125
Chief audit expert	96	Market operation technician	102	CTO	12
CIO	3	Innovation technician	32	Commercial bi insurance coordinator	82
Business specialist	114	Senior technician - office risk management	25	Head of engineering and databases	14
Retail risk modelling team leader	19	Senior marketing & communications	31	Financial risk analyst	
Head of it department	19	It manager	26	Investment and financial analyst	48
Senior mis analyst deputy manager	8	Innovation adviser	43	Compliance analyst	80
Research manager	26	Marketing adviser	84	System developer	52
Data governance manager	16	AML expert	56	ICT manager	19
Operations manager	60	Investing consultant	60	AI scientist	94
Audit expert	90	Accountant	70	Cloud systems expert	24

Table 2. List of selected domain knowledge experts

Before selecting any strategies or innovation lines in any organizations, it is better to design a comprehensive and in-depth survey tool. Having selected and rated the most important factors and their associated technologies that would help decision making core team to make the validations to come up with concrete results. In this way, the interactive and evolving questionnaire which includes the designed questions using the key concepts extracted at Phase 1 is the second task of Auditing phase. Furthermore, it includes digital and financial technologies and proposes know-how ideas about corresponding markets, their developments, and their relationships among each other. The questionnaire asking experts to determine the TRL maturity score of in-house or available technologies based on their experiences. Likert scale descried at [49] is used in this paper to shape the related survey questionnaire that will help to capture the view and insights of domain experts across different stages of the methodology. Table 3, summarizes the key factors extracted from Phase 1, and used to shape the questionnaires at Phase 2.

Improving customer experience and satisfaction	Developing a cloud strategy
Successful launching new products and services	Aligning IT/telecom with business strategies
Improving operational efficiencies	Ensuring regulatory compliance
Boosting creativity and innovation	Tactical solution to overcome overwhelming tasks due to volume of data
Enhancing sales and marketing effectiveness	Dealing with lack of standards and interoperability
Rapidly responding to market demand & disruption	Cybersecurity
Expanding to new markets, to new regions	Digital Marketing (Omnichannel Marketing, Marketing Automation)
Aligning vision and goals across organization	Data Analytics
Hiring and retaining talent	Data Centers and Cloud Infrastructure
Satisfying key stakeholders	Enterprise Resource Planning (ERP)
Cost savings	Digital Commerce
Customer retention rates	Enterprise Mobility Management
Customer acquisitions	Customer Experience Management
Speed to market	Enterprise Information Management (Enterprise Content Management)
Employee retention rates	Unified Communications & Collaboration (UCC)
Stock price increase	Robotic Process Automation
Enhanced Systems integration; improved managing multi-vendor solutions	Crowdfunding Platforms
Ensured network stability/reliability	Mobile Payments
Dealing with skills shortage; tailored training requirements	Robo-Advisors

Precise size estimations of financial investment or uncertain ROI	Blockchain
Dealing with privacy concerns	Computing
Reducing fear among users to adopt new technologies	Network communication systems

Table 3. Key factors to design questions for questionnaires

Phase 3; Evaluation and benchmarking phase: After collecting survey results and preparation of the data, a PCs analysis will be conducted on the data to visualize different dimensions of the multivariate data. The main feature of the proposed PCs is its uniqueness that is using an evolved PCA as a statistics-based factor weighing approach benefiting from its "comprehensive factors" selection algorithm. The selection of "comprehensive factors" are getting conducted autonomously by the paper's proposed PCs approach. Compared to Chachlakis et al. approach described on [50], the PCA algorithm embedded in the PCs tool selects the most correlated factors without compromising the order of adjacent features regarding the feature importance and multivariate Gaussian distribution-based correlation analysis. In order to reduce the clutter, it is necessary to have the least possible but optimized numbers of features to visualize correlations and interconnections among features.

The results of PCs analysis will be compared to tested hypotheses as the ground truth designed by roadmapping experts for the sake of the sanity check.

4 Results

In the previous section, we defined the necessary tools and techniques to execute the proposed methodology of producing roadmap scenarios those are combinations of factors extracted from explicit and tacit knowledge appeared on the domain's written literature. In addition, we emphasized the importance of sanity check and applying benchmarking on the results of technology roadmapping and decision making. In this way, we fulfilled the first two phases of the proposed methodology (phase 1. General data collection phase and phase 2. Auditing phase) to collect data, extract the explicit knowledge, shape the questionnaires, and collect the survey results. Then we shared the main concepts of digitalization extracted from the first phase (General data collection phase) of the methodology with the 7 focus groups those each consists of 400 experts, randomly assigned from the experts' list given on Table 2. Each focus group is supposed to design a list of hypotheses while they are shaping their proposed technology roadmap upon these hypotheses. After collecting all these roadmaps and hypotheses, we applied a two-stages results analysis to perform the benchmarking. **The first stage** is using statistical hypothesis test to accept or reject each hypothesis shaped by the focus groups. **The second stage**, is to probe the validity of the results at stage 1, using the proposed PCs tool.

4.1. Focus group hypotheses

As a use case for this paper, we are going to investigate deeply, the result issued and supported by one of the seven focus groups. Note that, we are going to apply hypothesis test that is independent from the use of PCs plot to visualize multivariate relations among different dimensions and features of a decision. At the end, the extracted results are going to be discussed, elaborated and compared to a ground truth roadmap (an already designed, approved and running roadmap in one of the involved banks).

Table 4, representing the roadmap designed by representatives of a focus group (one of the seven focus groups). They have defined targets and the necessary steps to fulfil these targets. In tis regards, they have used the corresponding hypotheses illustrated on the table. They have mainly emphasized the importance of "data science" (DS) and" computing systems" as well as "Blockchain technologies" to fulfil the digitalization journey that must result to numeric improvement on "customer churn", "customer satisfaction" and "marketing presence".

The target	Steps
80% and more efficiency on customer retention	Digital presence of the company needs to be improved above 80% of the most successful example in the market
80% and more efficiency on customer satisfaction	Efficiency of business models need to be improved by on-line banking activities
80% and more efficiency on marketing effectiveness	Improvement on reaching out to SMEs by the company
	Company needs to migrate to federated data systems and cloud
	Relying on cloud systems and DS will reduce capital investments on standardization practices
	New markets need to be recognized.
	New business models must be designed based on blockchain technologies
	Robo-advisory and face recognition technologies need to be implemented to the service platforms
Hypotheses	
	DS are cost-effective practices
	Success score of business ecosystem participation is correlated to the success of DS-based business models
	DS provide new markets
	SMEs can benefit of DS to catch up with new trends
	DS can make bridge among different technologies and solutions
	APIs are correlated to the successful DS models
	Being a node in a DS helps companies to have better access to customers and their data
	Higher chance of accessing to information in federated architecture using DS
	General standardizations are adequately well-designed in DS
	DS help companies to identify their role in the ecosystem in a proper level of granularity

Table 4. The steps of roadmap designed by representatives of a focus group

Concerning the significance of the hypotheses, we applied a “one-sample t-test” to evaluate the soundness of the designed hypotheses and eventually the proposed roadmap. Fig. 6, illustrates the results of one-sample test. These tests have been applied on the results from the survey data that is depicted on the normalized version of our survey results illustrated on Fig. 7. In order to be able to give a notion of our aggregated results, we needed to normalize (to make the real-data anonymous) our data for dissemination and publishing targets.

One Sample Test						
	Test Value = 50					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Development of companies is independent from DS experts	4140,00	399,00	0.000	277500,00	14533,00	40967,00
Business models' perfection are independent from DS	10367,00	399,00	**0.3005**	25,139.81	-22,533.7467	72,813.3667
Companies can build sustainable business models using DS	2505,00	399,00	0.0126	3577525,00	769852,00	6385198,00
Data sharing associates to the success of business models	23804,00	399,00	0.0178	3384475,00	589295,00	6179655,00
DS contribute to innovative business models	22545,00	399,00	0.0247	3190225,00	40837,00	597208,00
Business models based on DS aware of ecosystems issues	43232,00	399,00	0.0001	67975,00	361718,00	997782,00
DS are cost-effective practices	4678,00	399,00	3.44E-5	80325,00	455941,00	1150559,00
Success score of business ecosystem participation is correlated to the success of DS-based business models	20913,00	399,00	0.0431	191875,00	62963,00	3774537,00
DS provide new markets	21241,00	399,00	0.0401	195.8	9344,00	382256,00
SMEs can benefit of DS to catch up with new trends	67924,00	399,00	4.02E-11	195.76	1391009,00	2524191,00
DS can make bridge among different technologies and solutions	58989,00	399,00	7.81E-9	2669625,00	1779921,00	3559329,00
APIs are correlated to the successful DS models	24984,00	399,00	0.0129	742.84	1583296,00	1,327.3504
Being a node in a DS helps companies to have better access to customers and their data	24654,00	399,00	0.0141	7328025,00	1484507,00	1,317.1543
Higher chance of accessing to information in federated architecture using DS	24561,00	399,00	0.0145	730075,00	1457025,00	1,314.4475
General standardizations are adequately well-designed in DS	13981,00	399,00	**0.1629**	39,027.6975	-15,849.8257	93,905.2207
DS help companies to identify their role in the ecosystem in a proper level of granularity	24867,00	399,00	0.0133	3549675,00	743398,00	6355952,00

Fig. 6. Hypotheses related to the desired roadmap and One-Sample Statistics (N=400)

It is obvious that all hypotheses are significantly valid (statistically speaking) except two hypotheses. "Business model's perfection are independent from DS" and "General standardisation are adequately well-designed in DS" hypotheses are **rejected**. This evaluation has been performed under 95% of confidence interval.

	DKE	Sales & marketing effectiveness	Digital presence background	Improve customer experience	Decision making acceleration	Background to expand to new markets, to new regions	Boost creativity and innovation	Improve collaboration	Reduce operational costs	Beat or keep up with competitors	...	Employee retention rates	Stock price increase
0	1	0.874918	0.174695	0.524951	0.935856	0.541225	0.541916	0.722554	0.698780	0.736433	...	0.000000	0.000000
1	2	0.622255	0.000000	0.000000	0.525975	0.641546	0.642364	0.385418	0.000000	0.261881	...	0.512356	0.000000
2	3	0.168928	0.337299	0.675713	0.478986	0.870826	0.697549	0.871937	0.168650	0.000000	...	0.869333	0.524914
3	4	0.297601	0.297110	0.446402	0.644274	0.767069	0.153609	0.460828	0.445666	0.782800	...	0.306301	0.000000
4	5	0.000000	0.634209	0.000000	0.172868	0.654951	0.262315	0.000000	0.000000	0.668383	...	0.261531	0.394789
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2848	2849	0.000000	0.147163	0.147406	0.302770	0.759878	0.608678	0.000000	0.000000	0.310185	...	0.151715	0.458037
2849	2850	0.727773	0.181643	0.909717	0.983072	0.187584	0.939116	0.375646	0.908216	0.191431	...	0.000000	0.753808
2850	2851	0.700361	0.699206	0.350181	0.137632	0.180518	0.180749	0.180749	0.524405	0.368441	...	0.180209	0.544061
2851	2852	0.464579	0.618417	0.309719	0.199066	0.159660	0.639456	0.479592	0.463813	0.162935	...	0.000000	0.000000
2852	2853	0.000000	0.802770	0.321639	0.462348	0.000000	0.498049	0.664066	0.321108	0.000000	...	0.496562	0.166572

2853 rows × 64 columns

Fig. 7. A screenshot of "*Jupyter notebook*"[8] environment for standard normalized survey results dataset

4.2. Parallel Coordinates

A benchmarking practice considering generated hypotheses from one of seven focus groups is provided in this paper. Hence, after receiving the desired hypotheses and the corresponding roadmap from the focus group, we apply our PCs on the survey dataset. As it is shown on Fig. 8, we have adjusted the PCs upon the three main objectives (targets) by the focus group those were 1. Better efficiency on customer retention, 2. Higher efficiency on customer satisfaction and 3. Boosted marketing

[8] https://jupyter.org/

effectiveness. Three variables of “V1”, “V2” and “V3”, representing accordingly, “Sales & marketing effectiveness”, “Digital presence background” and “Improved customer experience”.

```python
fig = go.Figure(data=
    go.Parcoords(
        line = dict(color = df['ColorVal_5'],
                   colorscale = 'Electric',
                   showscale = True,
                   cmin = 11,
                   cmax = 43),
        dimensions = list([
            dict(range = [0,1],
                 constraintrange = [0,1],
                 label = "V1", values = df['Sales & marketing effectiveness']),
            dict(range = [0,1],
                 label = 'V2', values = df['Digital presence background']),
            dict(tickvals = [0,1],
                 label = 'V3', values = df['Improve customer experience']),
            dict(range = [0.2,0.4],
                 label = 'V4', values = df['Decision making acceleration']),
            dict(range = [0,1],
                 visible = True,
                 label = 'V5', values = df['Background to expand to new markets, to new regions']),
            dict(range = [0,1],
                 label = 'V6', values = df['Boost creativity and innovation']),
            dict(range = [0,1],
                 label = 'V7', values = df['Tactical solution to overcome overwhelming tasks due to volume of data']),
            dict(range = [0,1],
                 label = 'V8', values = df['Data Analytics']),
            dict(range = [0,1],
                 label = 'V9', values = df['Data Centers and Cloud Infrastructure']),
            dict(range = [0,1],
                 label = 'V10', values = df['Improve collaboration']),
            dict(range = [0,1],
                 label = 'V11', values = df['Reducing fear among users to adopt new technologies']),
            dict(range = [0,1],
                 label = 'V12', values = df['Aligning IT OR telecom with business strategies']),
            dict(range = [0,1],
                 label = 'V13', values = df['Ensuring regulatory compliance']),
            dict(range = [0,1],
                 label = 'V14', values = df['Cybersecurity'])])

    )
)
fig.show()
```

Fig. 8. Cropped screenshot from the developed python code depicting factors information.

According to PCs, it is possible to see the various relations within multivariant set of data and between different features. As shown on Fig. 9, the. traditional PCs tend to show all possible interconnected features for all exemplary 14 features.

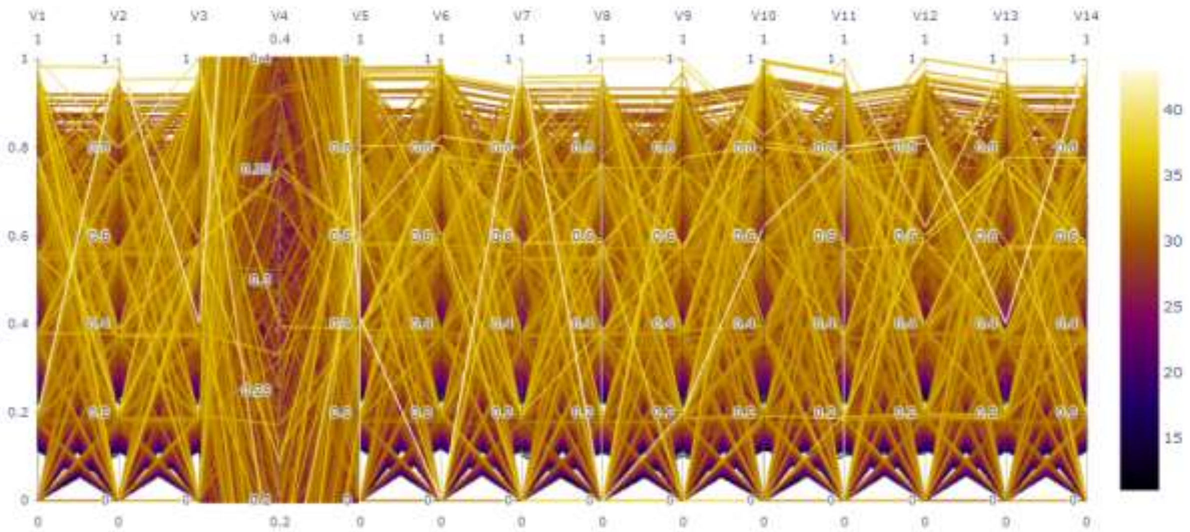


Fig. 9. Traditional PCs: all possible roadmaps consisted of all 14 features

However, our proposed PCs is using its unique data-based weighing system assigning and selecting the most effective features with dimension reduction and using various range of colour spectrums that is starting from the least importance (0=Black) to the most importance (50=Gold). As it is illustrated on Fig. 10, there are already thousands of roadmap scenarios (combinations of tasks) even tough if the system considers the ecosystem's 10 most comprehensive factors those are selected considering the **pre-defined targets** (features and ranges) as well as using the multivariate Gaussian distribution based PCA algorithm for correlation analysis.

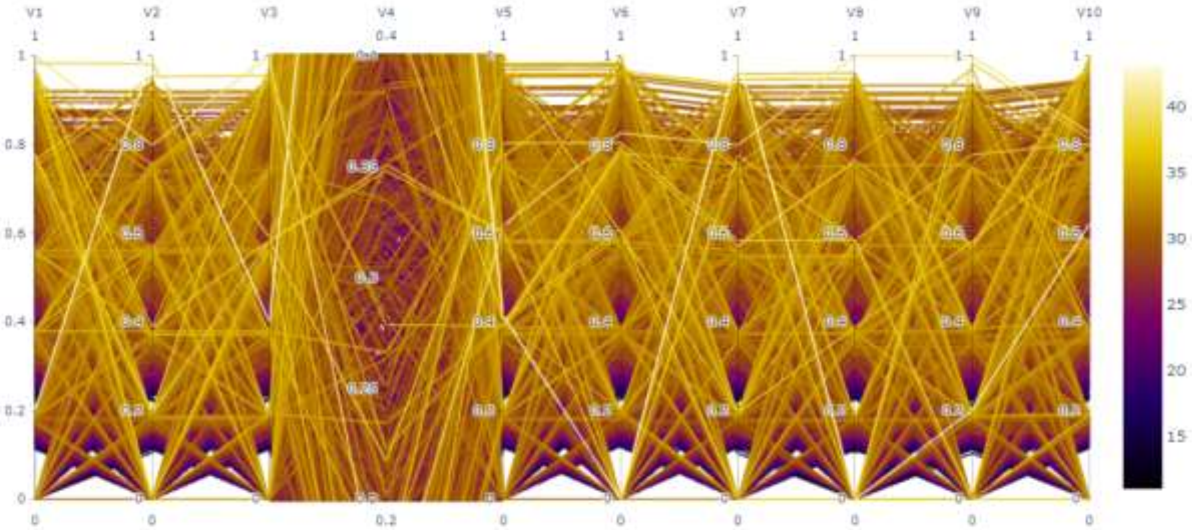


Fig. 10. Proposed PCs; all possible roadmaps consisted of 10 most important factors

As it is shown on Fig. 11, we applied the focus group's targets those were 80% and more efficiency on customer retention, 80% and more efficiency on customer satisfaction and 80% and more efficiency on marketing effectiveness those are represented by "Sales & marketing effectiveness =V1", "Digital presence background =V2" and "Improved customer experience =V3" (in pink colour).

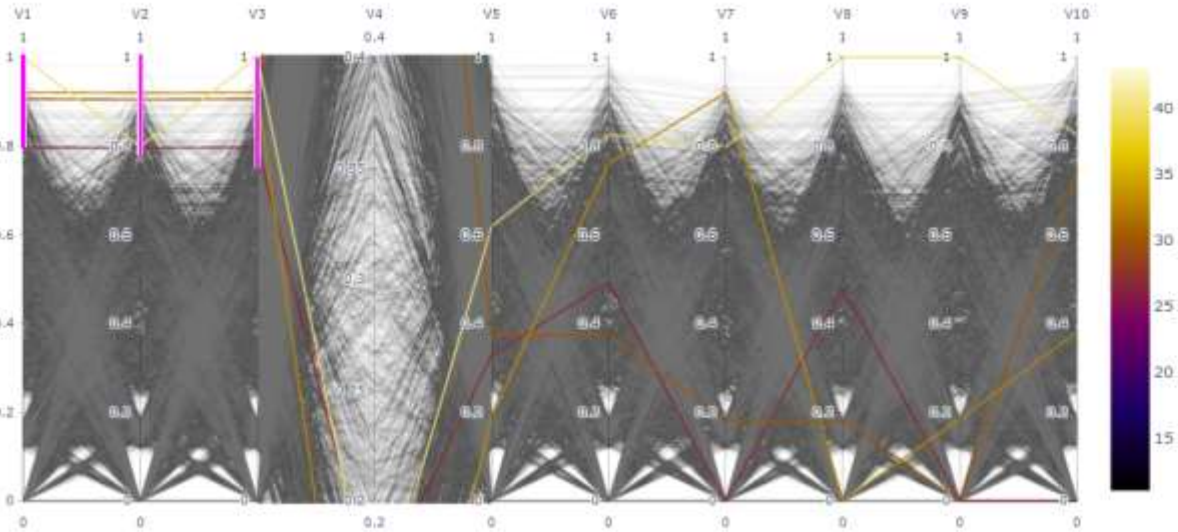


Fig. 11. PCs after applying the target measures

After applying the focus groups targets in the PCs, we observe on Fig. 12, an array of golden colour representing the selected roadmap that has considered "V6", "V7", "V8", "V9" and "V10" features as very important factors to realize the focus group's targets. It is interesting and valid according to the hypotheses test results, that "Tactical solution to overcome overwhelming tasks due to volume of data =V7", "Data Analytics =V8" and "Data Centres and Cloud Infrastructure =V9" factors are data-related tasks. Similarly, the other factors like "Boost creativity and innovation=V6" and

"Improve collaboration=V10" are highly connected to the data-related tasks and must be considered while developing business models. Briefly the provided roadmap by the PCs pushing the fact of focusing on data related technologies similar to the hypotheses provided by the focus group. However, even though Fig. 12 has provided lots of insights, the "'Decision making acceleration=V4" factor which is a very important feature is completely out of access and no golden coloured line has been appeared. This is due to the bad ordering at defining adjacent features that is one of the very common shortcomings of PCs plots. Thus, we activate the feature ordering function of our proposed PCs that delivers Fig. 13, with completely new ordering of adjacent features.

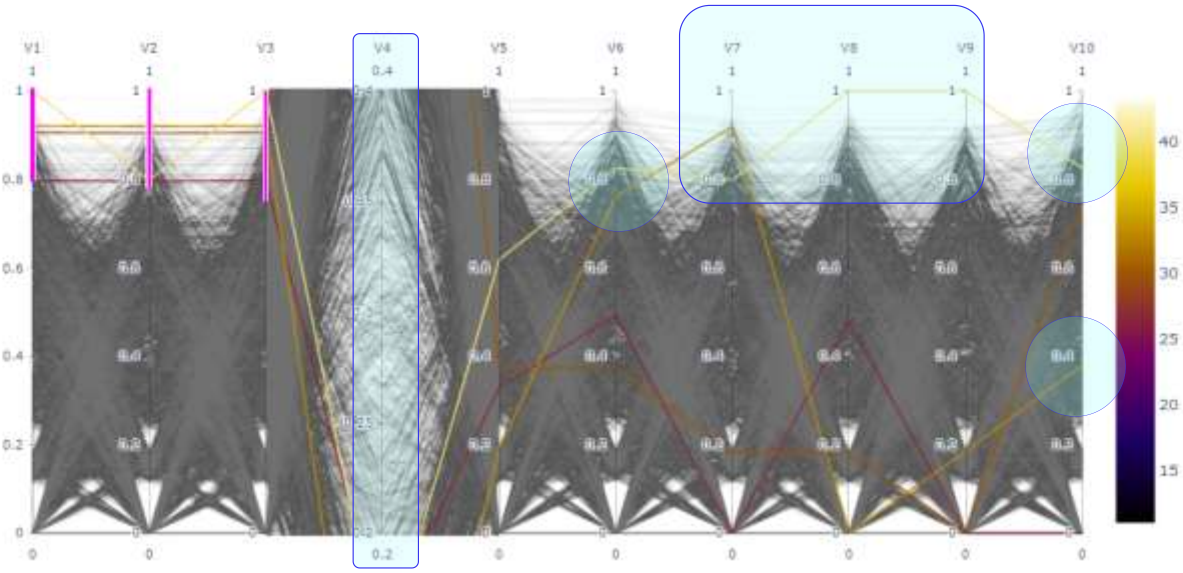


Fig. 12. The roadmap provided by the proposed PCs **without adjacent ordering**

According to the output plot after executing the adjacent re-ordering function, the exact value of "V4" factor which is between 0.25 and 0.3 is more obvious. This achievement also shows the correct orders of the needed steps to realize a digitalization roadmap.

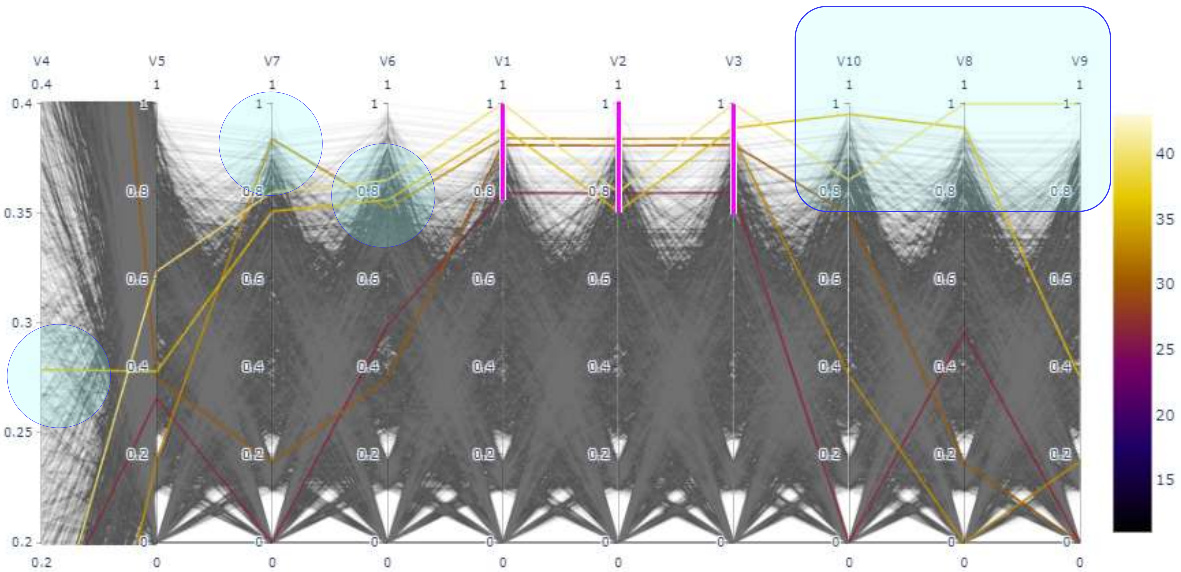


Fig. 13. The roadmap provided by the proposed PCs **with adjacent ordering**

However, according to Fig. 14, if we stuck with the traditional PCs results and try to interpret them in order to make a sound decision, we would make a definite mistake regarding our hypotheses results. On Fig. 14, it is obvious that “V7”, “V8” and “V9” have been impacted negatively due to noises from “V11”, “V12”, “V13” and “V14”. For example “Cybersecurity =V14” is of paramount importance factor for every digitalization journey at IT departments and must be considered as mandatory factor. In here, for the case of the focus group’s targets those were 80% and more efficiency on customer retention, 80% and more efficiency on customer satisfaction and 80% and more efficiency on marketing effectiveness, has no direct relation though.

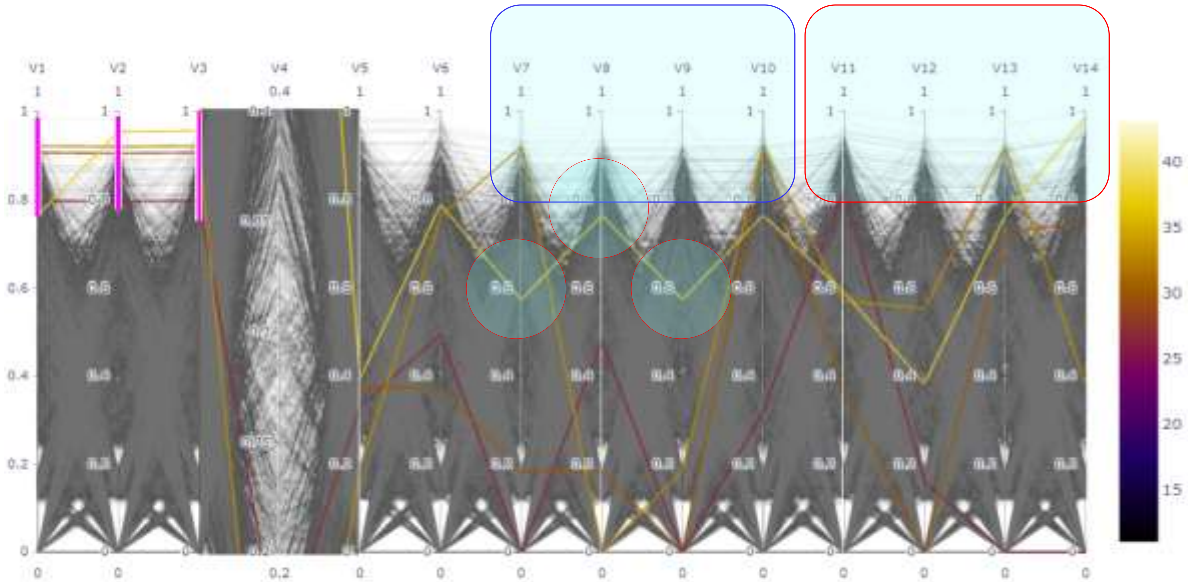


Fig. 14. The roadmaps provided by the traditional PCs without embedded PCA

5 Discussion

5.1. Discussion over the proposed methodology

Providing hypotheses and testing them, are two important tasks that mainly domain knowledge experts are fulfilling them in the industry. In addition, scholars from academia, are also active in this domain if they have access to the domain knowledge and real data. Admitting the fact that domain literature knowledge is as valuable as the knowledge of domain knowledge experts. Therefore, merging these two sources of knowledge could give an extra boost while decision makers are preparing their strategic roadmaps. Thus, the proposed methodology depicted an architecture to capture the most important features those are tacitly hidden in the contexts or reports. However, exploring knowledge from text could be much more difficult task from text scrapping. Specially when one is dealing with tacit knowledge. Applying the proposed methodology to the real-life data has revealed the challenges regarding the technical statements those are mainly covered and not explicitly described. One of these challenges is the lack of comprehensive lexicon and terms in the literature those are overlapping with the technical lexicon being used in the industry by domain knowledge experts. Another challenge is to have access to the solid DKE via surveys or other knowledge collection channels.

5.2. Discussion over the proposed Parallel Coordinates tool

Multivariate data with multiple dimensions are always complicated and confusing when it comes to visualize them. Generating hypotheses considering characteristics of variables as well as their relationships to each other would be a simplified task using an easy to interpret PCs tool. However, most of the times interpreting PCs is extremely difficult, specially when someone is dealing with a highly dimensioned dataset. clutter and false order of features (vertical lines) are the most important challenges that have cost the reputation of PCs. This shortcoming reveals even more when it comes to visualize texts and subject-matter features. However, despite these issues, PCs plot is one of the few tools that we can use to visualize multiple dimensions to do exploratory data analysis and compare the distributions and observing layered correlations. As it is illustrated on this paper, the application of PCA merged into a PCs' algorithm is highly effective to reduce clutter and expand the applicability of PCs while easing its interpretability. Another missing part in both PCA and PCs was their incapability to select the most important factors while a solution space of a feature (a range on a feature) is targeted. The proposed PCA algorithm has tackled this issue using the concept of integral in the PCA formulation. The interoperability and soundness of the technic is approved by implementing a benchmarking. In addition to handle the adjacent definition problem in traditional PCs, we proposed a built-in solution using PCA which functions according to request by the user. It navigates to propose the best order with changing the order of each adjacent using neighbourhood clusters up to 7-neighbour factors. However, this function needs large capacity of GPU capacity or HPC power to perform this functionality on larger number of adjacent factors.

5.3. Discussion over the use case and its results using the methodology

Our use case study was chosen to follow three targets: 1. A use case to validate the proposed methodology, 2. A use case to confirm the applicability and soundness of the proposed PCs and 3. A self-standing project to practice valid decision making and technology roadmapping for digitalization in finance industry. As a matter of fact, the first two have been discussed duly at the previous sections and it is worth it to discuss the driven insights over the generated and tested hypotheses as listed below:

1. One important dimension of technology roadmapping is to align the understanding of the current position of the company, its overall and strategic objectives beyond digitalization. Thus, in line with the technology watch, the technology audit and the background analysis mentioned for the Phase 1 implies to define indicators and metrics (whether qualitative or quantitative) to set the strategic goals of digital transformation in order to evaluate their achievement in the latter stages of development. This is also fundamental to enable an algorithmic process of decision-making regarding the digital transformation. These indicators have to take into account the context of the company, considering its financial capabilities, and providing measures to consider (if possible compute) ROI Fig.s.

2. Investigating the trend of global technologies and performing surveys to inquire available and on-going research, could be of great help to establish a correct methodology of technology evaluation. In this regards, technologies and their variants such as Cybersecurity, Digital Marketing (Omnichannel Marketing, Marketing Automation), Data Analytics, Data Centers and Cloud Infrastructure, Enterprise Resource Planning (ERP), Digital Commerce, Enterprise Mobility Management, Customer Experience Management, Enterprise Information Management (Enterprise Content Management), Unified Communications & Collaboration (UCC), Robotic Process Automation, Crowdfunding Platforms, Mobile Payments, Robo-Advisors, Blockchain, Computing, and last but not least, Network communication systems could be evaluated and selected from organizational needs perspective.
3. An important factor is concerning the required skills, not only IT but also domain knowledge, and this may also be disrupted by the digital transformation. It implies also assessing what kinds of internal processes, how the current products or services, how the relationships with market, could be impacted in a broad way.
4. Proper measuring of digital transformation impacts within and between organizational technologies and departments, will pave the way to reach an appropriate way of scalable continuum of resources evaluation through IT drivers. Concerning finance industry, the most related audit questions among inter and within organizational levels are:
    a. The effectiveness of sales & marketing department
    b. The projected background of digital presence
    c. Improvement percentage of reported customer experience
    d. Acceleration rate at decision making processes (do we need robotic process automation?)
    e. Background capability/potential to expand to new markets, to new regions
    f. The internal and outsourced abilities around creativity and innovation
    g. Improved levels of collaboration and management

    h. Possibility of reducing operational/capital investment costs using the emerging technologies
    i. Beat or keep up with competitors
    j. Attract and retain skilled workforce or establish proper timelines to re-skill and re-train the available workforce

5. It is of paramount importance to evaluate the relative position of the company pertaining to the state-of-the-art technologies (relation with phase 1), and to score its position against its competitors. This can be done using a digital maturity model, to draw the pre-digitalization landscape of the company, to assess objectively the used technologies, but also if they are combined in the right way for the relevant digital transformation assets considered.

6. It is also important to evaluate the current state of legacy systems and how they might be impacted by the intended transformation. Dealing with legacy systems is a strong issue related to digital transformation, leading to possibly costly and risky changes to the organization, especially as they have direct and indirect implications. They are built in accordance with the needs of specific business models, which are still relevant for some of them, they required important expenses that must be amortized, they gave rise to a specific working culture that has to evolve. At the same time, their adaptation is leading to costs that are not to be borne by newcomers in the market, leading to potential competitiveness issues.

7. Mainly, the methodologies that made up the financial technology development roadmap process were showing diverse complexities from one organization to another. The literature on road maps focusing on the relationship between financial technology and product, where the life cycles of the product or service shorten and become more diverse.

8. It requires determining what part of the business model may be impacted by the intended transformation. For this, it may be convenient to re-use the business model canvas as any of its component may be disrupted. Digital transformation has indeed implications for value generation or the channels. The relationships with customers (and providers as well) may be affected. The background includes the analysis of more intangible components, including components such as the internal working culture.

Conclusion

The proposed approach introduces a new application of parallel coordinates plots to map consensus variabilities in a sound way regardless of their complexity, subjectiveness, and implicitness that is the nature of multivariate data to compose and test digitalization roadmapping hypotheses. The advanced parallel coordinates visualization tool provides autonomous feature selection and adjacent re-ordering function to tackle clutter and adjacent ordering issues of the current parallel coordinate's plots. Furthermore, it shows a unique application of single valued neutrosophic sets in tacit knowledge extraction from web scraped data and research documents to help shaping survey questionnaires and hypotheses. Also, the paper proposed a new approach to progress with strategic decision-making challenges in finance industry using anonymised real-life data. The data is collected from pilot execution on a cluster of a banking industry during a research project that is mentioned at acknowledgment. The data is now open access at GitHub[9]. The superiority of the proposed methodology and the sanity of the used methods are approved after a benchmarking practices. The proposed roadmaps by the methodology have been validated by all focus groups. However, the necessity of a partial human interaction is proposed by DKEs to maximize the functionality of the approach. Thus, they considered this approach as "semi-autonomous" assisting technology for roadmapping activities. For future work,

[9] https://github.com/PeimanAlipourSarvari

we are going to propose a neutrosophic fuzzy sets approach to enhance the functionality of feature re-ordering and adjacent selection.

Data availability statement
The used dataset is open access to public on GitHub under the https://github.com/PeimanAlipourSarvari/data_strategic_roadmap_fintech link.

Declaration of Competing Interest
The authors declare that they have no conflict of interest regarding the publication of this article.

Acknowledgments
This research is financed by the European Union's Horizon 2020 research and innovation programme under grant agreement No 810318 (Project "Industry 4.0 impact on management practices and economics (IN4ACT)".

References
1. Hair, J. F., Page, M., & Brunsveld, N. (2019). *Essentials of business research methods*. Routledge.

2. Daft, R. L. (2015). *Management*. Cengage Learning.

3. Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., & Ohlson, D. (2012). *Structured decision making: a practical guide to environmental management choices*. John Wiley & Sons.

4. Hammond, J. S., Keeney, R. L., & Raiffa, H. (2015). *Smart choices: A practical guide to making better decisions*. Harvard Business Review Press.

5. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018, October). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)* (pp. 80-89). IEEE.

6. Ghobakhloo, M. (2018). The future of manufacturing industry: a strategic roadmap toward Industry 4.0. *Journal of manufacturing technology management*.

7. Dwivedi, Y. K., Hughes, D. L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S., ... & Upadhyay, N. (2020). Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life. *International journal of information management*, *55*, 102211.

8. Liberman-Yaconi, L., Hooper, T., & Hutchings, K. (2010). Toward a model of understanding strategic decision-making in micro-firms: exploring the Australian information technology sector. *Journal of Small Business Management*, *48*(1), 70-95.

9. Cenamor, J. (2021). Complementor competitive advantage: A framework for strategic decisions. *Journal of Business Research*, *122*, 335-343.

10. Kane, G. C., Phillips, A. N., Copulsky, J., & Andrus, G. (2019). How digital leadership is (n't) different. *MIT Sloan Management Review*, *60*(3), 34-39.

11. Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. *Information Systems and e-Business Management*, *16*(3), 479-491.

12. Priyono, A., Moin, A., & Putri, V. N. A. O. (2020). Identifying digital transformation paths in the business model of SMEs during the COVID-19 pandemic. *Journal of Open Innovation: Technology, Market, and Complexity*, *6*(4), 104.

13. Huang, M. H., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research*, *24*(1), 30-41.

14. FitzGerald, J. D., Dalbeth, N., Mikuls, T., Brignardello-Petersen, R., Guyatt, G., Abeles, A. M., ... & Neogi, T. (2020). 2020 American College of Rheumatology guideline for the management of gout. *Arthritis Care & Research*, *72*(6), 744-760.

15. Alipour Sarvari, P., Nozari, M., & Khadraoui, D. (2019). The potential of data analytics in disaster management. In Industrial engineering in the big data era (pp. 335-348). Springer, Cham.

16. Sarvari, P. A., Ustundag, A., Cevikcan, E., Kaya, I., & Cebi, S. (2018). Technology roadmap for Industry 4.0. In Industry 4.0: Managing the digital transformation (pp. 95-103). Springer, Cham.

17. Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., ... & Erickson, T. (2019, May). How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-15).

18. Srinivasa-Desikan, B. (2018). *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd.

19. Che, T., Wu, Z., Wang, Y., & Yang, R. (2018). Impacts of knowledge sourcing on employee innovation: the moderating effect of information transparency. *Journal of Knowledge Management*.

20. Moffitt, K. C., Rozario, A. M., & Vasarhelyi, M. A. (2018). Robotic process automation for auditing. *Journal of emerging technologies in accounting*, *15*(1), 1-10.

21. Ritchie, T. J., Ertl, P., & Lewis, R. (2011). The graphical representation of ADME-related molecule properties for medicinal chemists. *Drug Discovery Today*, *16*(1-2), 65-72.

22. Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent advances and challenges. *The Visual Computer*, *30*(12), 1373-1393.

23. Ermentrout, B., & Mahajan, A. (2003). Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students. *Appl. Mech. Rev.*, *56*(4), B53-B53.

24. Bertini, E., Tatu, A., & Keim, D. (2011). Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2203-2212.

25. Alminagorta, O., Loewen, C. J., de Kerckhove, D. T., Jackson, D. A., & Chu, C. (2021). Exploratory analysis of multivariate data: Applications of parallel coordinates in ecology. Ecological Informatics, 64, 101361.

26. Baskurt, G., Martin, S. A., Sarvari, P. A., & Khadraoui, D. (2019, September). Open Data Availability and Suitability for Financial Analyses. In *Global Joint Conference on Industrial Engineering and Its Application Areas* (pp. 279-290). Springer, Cham.

27. Alipour Sarvari, P., Martin, S. A., Baskurt, G., Nozari, M., & Khadraoui, D. (2020, August). Regenerative Supply Chain Through Digitalization in Diary. In *Global Joint Conference on Industrial Engineering and Its Application Areas* (pp. 377-389). Springer, Cham.

28. Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, *1*(1), 100008.

29. Hartmann, J., & Van Keuren, L. (2019). Text mining for clinical support. *Journal of the Medical Library Association: JMLA*, *107*(4), 603.

30. Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. *Electronics*, *11*(3), 374.

31. Michael, W. B., & Berry, M. C. (2003). Survey of text mining: clustering, classification, and retrieval. Amazon.

32. Feng, L., Chiam, Y. K., & Lo, S. K. (2017, December). Text-mining techniques and tools for systematic literature reviews: A systematic literature review. In 2017 24th Asia-Pacific Software Engineering Conference (APSEC) (pp. 41-50). IEEE.

33. Ignatow, G., & Mihalcea, R. (2017). An introduction to text mining: Research design, data collection, and analysis. Sage Publications.

34. Bounabi, M., Elmoutaouakil, K., & Satori, K. (2021). A new neutrosophic TF-IDF term weighting for text mining tasks: text classification use case. International Journal of Web Information Systems.

35. Brooking, A. (1998). Corporate memory: Strategies for knowledge management. International Thomson Publishing.

36. Polanyi, M. (1966). 1983. The tacit dimension. Gloucester, MA: Peter Smith. Orign. Pub

37. Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. MIS quarterly, 107-136.

38. Nonaka, I. (2009). The knowledge-creating company. In The economic impact of knowledge (pp. 175-187). Routledge.

39. Gordon, J. E. (2006). The new science of strong materials: or why you don't fall through the floor (Vol. 27). Princeton University Press.

40. Davies, M. (2015). Knowledge–Explicit, implicit and tacit: Philosophical aspects. International encyclopedia of the social & behavioral sciences, 13, 74-90.

41. Heinrich, J., & Weiskopf, D. (2013). State of the Art of parallel coordinates. *Eurographics (State of the Art Reports)*, 95-116.

42. Lind, M., Johansson, J., & Cooper, M. (2009, July). Many-to-many relational parallel coordinates displays. In *2009 13th International Conference Information Visualisation* (pp. 25-31). IEEE.

43. Heinrich, J., & Weiskopf, D. (2015). parallel coordinates for multidimensional data visualization: Basic concepts. *Computing in Science & Engineering*, *17*(03), 70-76.

44. Mitku, A. A., Zewotir, T., North, D., & Naidoo, R. N. (2020). Exploratory data analysis of adverse birth outcomes and exposure to oxides of nitrogen using interactive parallel coordinates plot technique. *Scientific reports*, *10*(1), 1-9.

45. Omarini, A. (2017). The digital transformation in banking and the role of FinTechs in the new financial intermediation scenario.

46. Piao, S., Ananiadou, S., Tsuruoka, Y., Sasaki, Y., & McNaught, J. (2007, January). Mining opinion polarity relations of citations. In International workshop on computational semantics (IWCS) (pp. 366-371).

47. Bounabi, M., Elmoutaouakil, K., & Satori, K. (2021). A new neutrosophic TF-IDF term weighting for text mining tasks: text classification use case. International Journal of Web Information Systems.

48. Majumdar, P. (2015). Neutrosophic sets and its applications to decision making. In Computational intelligence for big data analysis (pp. 97-115). Springer, Cham.

49. Derrick, B; White, P (2017). "Comparing Two Samples from an Individual Likert Question". International Journal of Mathematics and Statistics. 18 (3): 1–13.

50. Chachlakis, D. G., Prater-Bennette, A., & Markopoulos, P. P. (2019). L1-norm Tucker tensor decomposition. IEEE Access, 7, 178454-178465.