Article

A Systematic Review of Recommendations of Long-Term Strategies for Researchers Using Data Science Techniques

Gilberto Ayala-Bastidas [ID], Hector G. Ceballos *[ID] and Francisco J. Cantu-Ortiz [ID]

Tecnologico de Monterrey, Monterrey 64849, Mexico; ab.gilberto@gmail.com (G.A.-B.); fcantu@tec.mx (F.J.C.-O.)
* Correspondence: ceballos@tec.mx

Abstract: The impact of the strategies that researchers follow to publish or produce scientific content can have a long-term impact. Identifying which strategies are most influential in the future has been attracting increasing attention in the literature. In this study, we present a systematic review of recommendations of long-term strategies in research analytics and their implementation methodologies. The objective is to present an overview from 2002 to 2018 on the development of this topic, including trends, and addressed contexts. The central objective is to identify data-oriented approaches to learn long-term research strategies, especially in process mining. We followed a protocol for systematic reviews for the engineering area in a structured and respectful manner. The results show the need for studies that generate more specific recommendations based on data mining. This outcome leaves open research opportunities from two particular perspectives—applying methodologies involving process mining for the context of research analytics and the feasibility study on long-term strategies using data science techniques.

Keywords: researcher career; data science; long-term strategies

1. Introduction

Sánchez-Torres and Miles described tools for a systematic assessment of the challenges and opportunities using future-oriented technology analysis to examine their application for the development of public policies in the context of e-government [1]. However, the existing reviews are distributed in different contexts, mainly focusing on medicine. These reviews do not present data science as a tool for generating strategies, but rather the strategies are part of empirical conclusions [2–7]. Suppose that, in this study, we add scientometrics or research analytics; in that case, the detection of this type of work becomes void. As far as we know, there are no comprehensive studies that collect, characterize and contextualize the case studies combining methodologies and tools. Hence, there is no systematic review focusing on long-term strategies within this context.

The purpose of this systematic review is to identify whether there are articles that present long-term strategies with data science, specifically process mining, within the context of research analytics. We aim to find long-term strategies and the methodologies used for developing and evaluating them. The interest lies in reviewing how strategies are described in the literature and the effects of their application in the long-term.

Gómez et al. [8] discussed that a systematic review of the literature offers a tool for identifying, evaluating and interpreting the currently published studies whose context is mainly defined. The systematic review makes it possible to synthesize the information rigorously and impartially so that what is proposed has a high scientific value. They also added that one of the main motivations for conducting a systematic review is detecting results for a particular topic of interest that might not otherwise be possible.

In this study, the research questions proposed are as follows.

RQ1: Are there methods of recommending long-term strategies using data science techniques?

RQ2: Are there studies with the above characteristics that use process mining within the context of research analytics?

The following are the objectives of this review study:

- Identify the main learning methods for long-term strategies.
- Identify and describe cases where recommendations for long-term strategies are made using data science.
- Identify and describe case studies where process mining has been used as a learning tool for long-term strategies.
- Identify and describe case studies where process mining has been applied in research analytics.

2. Background

2.1. Research Analytics

Scientometrics is the study and analysis of science through the calculation of indicators by data analysis methods. These indicators are commonly applied in science, technology and innovation. Leydesdorff and Milojevic' [9] provided the state-of-the-art scientometric development and evolution as a discipline in charge of the quantitative study of science, which began with Eugene Garfield's idea of an index to improve information retrieval. Cantu et al. [10] described research analytics as a set of techniques that allow information to be collected from large datasets to help organizations make decisions through statistics and probabilistic methods. Data analytics includes statistical techniques and involves methods from different fields, such as automatic learning, artificial intelligence, data science, and data mining. All these tools offer a range of possibilities for processing large amounts of data. The objective of its application lies in hypothesis testing, discovery of information and decision-making.

2.2. Data Science

Data science is an area that composes a set of techniques and methodologies whose main objective is to manipulate and process data to achieve retrieval of knowledge. Some of the activities in data science for decision-making and information prediction are extraction, preparation, exploration, transformation, storage, recovery and infrastructure, and mining [11]. On the one hand, Van der Aalst [12] defined it as an interdisciplinary area whose purpose is to convert data into a real value. On the other hand, Waller and Fawcett [13] described it as applying qualitative and quantitative methods, allowing prediction outcomes during problem resolution. Provost and Fawcett [14] also defined it as a set of fundamental principles that allow information to be extracted from a set of data and translated into knowledge. These principles involve techniques and processes that make it possible to understand a phenomenon by analyzing data to improve decision-making. The value can be described in terms of feasible information for decision-making, visualization, or any valuable insights resulting from the application. Data science also tries to take into account the ethical, social and business aspects. For example, in [15], the term "Green data science" is coined to refer to data science that makes responsible use of information in any of its dimensions in terms of the amount of data collected, where information is protected from misuse or irresponsible use.

2.3. Process Mining

Process mining is an area of research that aims to extract knowledge through event logs. It involves using event logs to automatically extract process-related information, as described in [16]. The event logs can be obtained from multiple and diverse sources. This is because of the large number of records made daily, from sending an email and registering in a new service to sharing an element in social networks. Each of these actions can be recorded as an event, which can be extracted for analysis. There are three main elements in process mining—discovery, conformance checking and enhancement. Discovery involves finding a process model from an event log. Conformance checking

allows verifying whether a proposed model corresponds to the actions being carried out and recorded in the event log. In performing this action, an event log is compared against a model of the same process to identify deviations, bottlenecks, or activities that may be conflicts within a process. Enhancement allows improving a previous process model by repairing or extending it. This implies implementing model improvements once the variations between the model and the event log are identified, providing a more realistic view of the analyzed process.

Depending on the event log characteristics, process mining can extract information from three different perspectives—control flow, organizational and case. For example, suppose that the event log describes the tasks performed in a certain process, and from these tasks, an ordering of activities can be inferred. In that case, the flow control perspective can be extracted. Suppose that the event log contains information that relates a certain activity to a particular resource (i.e., who performs that activity). In that case, the organizational perspective can be mined. This implies that social networks can be built and identify the relationships between resources and activities. Suppose that the event log has details of the tasks that can provide information on what influences one activity. In that case, the case perspective can be mined, allowing predictions to be made on the basis of the previously analyzed cases. Based on the aforementioned information, it can be deduced that each of the perspectives tries to answer the following—how? (control flow perspective), who? (organizational perspective), and what? (case perspective) [17].

Other techniques have been used to analyze sequences of events. Some approaches include neural networks and hidden Markov models [18,19]. However , one of the main disadvantages of neural networks is mainly because the resulting model is illegible to humans due to no process model from the point of view of the flow of activities (e.g., as can be seen in a Petri Net or a Business Process Model and Notation BPMN diagram).

The hidden Markov models are an extension of a Markov process composed of states and transition probabilities. Unlike the traditional model, there is a possible observation in each state, which has a probability; however, the state remains hidden. They are at a deeper level of abstraction from other systems of process representation. When using hidden Markov models, some questions remain to be answered (i.e., how to calculate the most transited path in a hidden Markov model or, given a set of sequences, how to find the hidden Markov model that maximizes the probability of producing those sequences).

Execution time of the executed iterative procedures is the main problem of using hidden Markov models. Finding the most transited path is one of the problems mentioned above, which belongs to the process mining domain. This problem can be solved with the Baum—Welch iterative algorithm; however, it is important to have a limited number of states. This feature also forms part of the challenges of the HMM application, which is the need to guess an appropriate number of states as input for the aforementioned algorithm. Like neural networks, the resulting model is inaccessible and difficult to interpret for the end-user, even in small examples. Although different data mining techniques can be exploited for process mining, they cannot perform the most relevant process mining tasks, such as process discovery, process enhancement and conformance checking [16].

3. Materials and Methods

Biolchini et al. [20] described the concept of a systematic review as a research methodology that seeks to analyze a predetermined topic by addressing it in a summarized, concrete, structured and reproducible manner [21]. Recently, a systematic review has been encouraged before the development of new projects [22], and this kind of publication seems to be increasing [23]. Although they have been criticized as secondary studies [24,25], they have also been qualified as a "fundamental scientific activity" [26].

One of the principal utilities of a systematic review lies in the examination of relevant studies through structured methods, as well as the identification of methodological inconsistencies or errors [26]. They also provide information on research gaps and provide a future vision to guide knowledge development [27]. Their main contribution is the ease

of identifying all studies on a selected topic; a lack of reviews would hinder this valuable scientific process by deliberately consolidating the relevant literature. This contribution is one of the main points with respect, for example, to literature reviews, which tend to be more general and may cover several aspects of the same topic. The main differences lie in the lack of assurance of a systematic protocol for obtaining and interpreting the results [28]. The results of systematic reviews, compared to literature reviews, may be perceived as more reliable [24,29,30]. Petticrew and Roberts [31] presented the processes in such a way to reduce biases while solving specific research questions. We took the most relevant characteristics outlined in the following from both approaches.

The following template is an adaptation of one of the aforementioned studies, including some elements present in the original proposal.

Protocol development. This stage involves the formulation of research questions (including focus, breadth, and quality), the definition of research objectives, and what is expected to be answered at the end of the systematic review. The identification and selection of research sources and the definition from which sources will be analyzed will be obtained.

Extraction of information. This involves the definition of inclusion criteria for evaluating the information obtained from the studies and determining its relevance and the definition of templates for extracting information and executing the extraction.

Presentation of results. This involves the presentation of trends and relevant classifications.

3.1. Question Formularization

Question focus. We identify methods of recommending long-term strategies and their relationship to data science to analyze whether there are coincidences with process mining within the context of research analytics.

Question quality and amplitude. This stage includes the following.

- Problem. Long-term recommendation strategies involve a set of activities that influence a future end result.
- Question. $\mu 0$: Which methods have been used for recommending long-term strategies within the context of research analytics?
- Effect. Identification of methodologies of recommendation strategies and matching with techniques specific to the area of data science.
- Outcome measure. The number of methodologies identified.
- Population. Publications related to long-term strategy recommendations, process mining, and research analytics.
- Application. Institutions that use data to calculate scientometric indicators for creating action plans and developing strategies for scientific production.

3.2. Sources Selection

- Source selection criteria definition. Publications currently available on the websites; presence of search engines using keywords; media and articles suggested by the experts.
- Studies language(s). English.
- Source identification. Source search methods: research through the web. Search string: different combinations of the following keywords were used for article retrievals, such as process mining, research analytics, long-term strategies, scientometrics, long-term learning, model, scientific career, and data mining. Table 1 describes the combinations, such as search strings.
- Source. Scopus.
- Sources selection after evaluation. The selected source meets the quality criteria.
- References checking. The source was approved by two researchers from the *Instituto Tecnológico y de Estudios Superiores de Monterrey*, who approved the use of the source in a consensual manner, according to the selection criteria.

Table 1. Query strings.

Query Number	Query String	Results	Included
1	TITLE-ABS-KEY ("process mining" research AND analytics)	37	29
2	"process mining" AND scientometrics	20	4
3	"long term strategies" AND scientometrics	12	4
4	"process mining" AND "long term learning"	1	0
5	TITLE-ABS-KEY ("scientometrics" AND "data mining")	48	27
6	"long term strategies" AND "data science" AND research AND analytics	1	0
7	"long term strategies" AND "data science"	4	1
8	TITLE-ABS-KEY ("scientific career" AND "researchers" AND model)	18	2
Total		141	67

3.3. Studies Selection

- Studies definition. In the studies' inclusion and exclusion criteria definition, we describe four IC inclusion criteria and two EC exclusion criteria defined in Table 2. In the procedures for study selection, as part of the procedure for selecting articles, the main criteria involve revising the document title, the abstract, and skimming of the article; in some cases, the complete reading of the article was required to achieve greater refinement. Study types definition—any study related to the research topic will be a candidate for selection.
- Selection execution. Initial studies' selection: 114. Selection review: 67.

3.4. Information Extraction

- Information inclusion and exclusion criteria definition. The information obtained from the studies should contain long-term techniques, methods, strategies, and use data science tools, if possible. Table 3 describes the criteria for the inclusion of information.
- Data extraction forms. The templates are defined for the identification of information for later use in the summary of results. They comprise author(s), title, year, abstract, author keywords, document type, and source.
  - Category: Association for Computing Machinery (ACM) taxonomy and ACM second-level taxonomy classification.
  - Subdomain: ACM third-level taxonomy classification.
  - Strategy application: a mention about strategies (Y/N).
  - Strategy purpose: how the strategy is applied, open.
  - Short-term strategy mention (Y/N): there is a mention of strategies in the short-term.
  - Long-term strategy mention (Y/N): there is a mention of strategies in the long-term.
  - PM: the methodology uses process mining techniques (Y/N): the most suitable process mining technique used in the experiments.

The eight basic queries in Table 1 can be concatenated into a single one, as shown in the following. TITLE-ABS-KEY ("process mining" research AND analytics) or ("process mining" AND scientometrics) OR ("long term strategies" , AND scientometrics) OR ("process mining" AND "long term learning") OR (TITLE-ABS-KEY ("scientometrics" AND "data mining")) OR ("long term strategies" AND "data science" AND research AND analytics) OR ("long term strategies" AND "data science") OR (TITLE-ABS-KEY ("scientific career" AND "researchers" AND model))

Table 2. Inclusion and exclusion criteria.

Criteria	Description
CI1	Includes publications whose titles contain long-term strategies, process mining, scientometrics, long term strategies, scientific career or recommendation.
CI2	Include publications that contain keywords that match the selected keywords.
CI3	Includes publications where the abstract contains selected keywords or whose theme indicates a visible relationship with the selected theme.
CI4	Includes publications that are available in full-text.
CE1	Excludes publications that do not meet the inclusion criteria described above.
CE2	Excludes any duplicate publication.

Table 3. Criteria for the inclusion of information.

Criteria	Description
IC1inf	Gather information about long-term strategies that have been used in any context.
IC2inf	Collect information on the contexts behind the implementation of long-term strategy recommendations.
IC3inf	Identify the models, methods or standards used in the implementation of long-term strategies.
IC4inf	Exclude information that is not related to the previously defined inclusion criteria.

4. Results

4.1. Study Selection

A total of 141 results were obtained from all queries (source: Scopus database, for all cases). Table 1 shows the total results per query. Those combinations that obtained 0 or no results were removed from the table. The inclusion and exclusion criteria described in Table 2 were applied to the 141 results. The inclusion criteria (Figure 1) were defined in the following order to determine whether a result should be included.

1. Identify if the result contains a title with any of the following words: process mining, scientometrics, long-term strategies, scientific career, or recommendation.
2. Verify that the article contains some of the keywords used to formulate queries.
3. Verify that the abstract contains some of the keywords used to formulate queries.
4. Identify if the publication is available or accessible in an open manner or through any of the Tecnologico de Monterrey's subscriptions.
5. The decision according to the following cases.

    (a) If at least one of the four criteria is met, the article is automatically included.
    (b) If none of the criteria are met, the item is discarded.

A total of 67 documents were obtained by applying these criteria. These documents were checked for the inclusion criteria defined in Table 3.

4.2. Study Characteristics

As a first step, each article was classified on the basis of the 2012 ACM Computing Classification System [32] taxonomy. We performed the following for the selection of a category within the taxonomy.

1. Search by keywords if there is a particular category to associate an article with.
2. Search the title for keywords to associate.
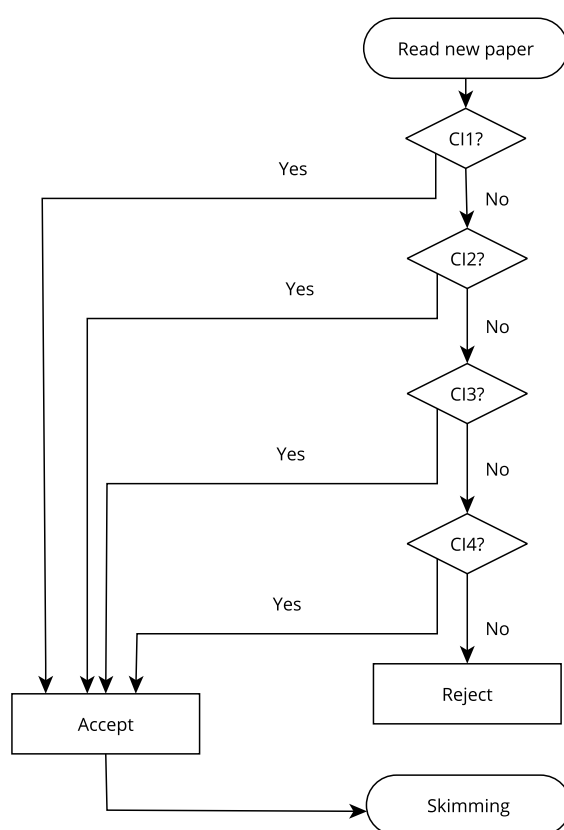3. Search for the abstract keywords to associate.

Figure 1. Flow diagram for inclusion criteria.

In some cases, taxonomy involved up to four levels of specificity; thus, the levels were defined as follows:

- level 1: category
- level 2: domain
- level 3: subdomain

Although there are more than four taxonomy levels, no more than three levels were necessary for this exercise. Those elements classified within one of the first three levels were labeled as not applicable in the subsequent levels. Table 4 shows the results for three levels (level 1: category (bold); level 2: domain (underline); and level 3: subdomain (italics)).

Naming the main topics is one of the main differences when classifying by ACM taxonomy. The results can be seen from the perspective described in Figure 2, depending on the keywords chosen. In the image, the main results correspond to three particular topics (scientometrics, data mining, and long-term strategies), implying that the study focuses on the relationship between these issues and that process mining is a key part of data mining. These results can be seen in Table 1, which shows the keyword combinations that obtained the greatest number of results. Figure 3 shows a bar chart of the proposed taxonomy. In it, we can see on the x-axis each of the main categories by level. Each category includes certain domains represented by each bar's colors according to the number of items present from each domain in that category. From here, we can see that the main category is related to Information systems, from which most of the papers are related to Data mining, followed by Search engine architectures and scalability and Retrieval tasks and goals. After this, Applied computing is the second most representative category, and Social and professional topics are perceived as the least representative category.

Table 4. Taxonomy.

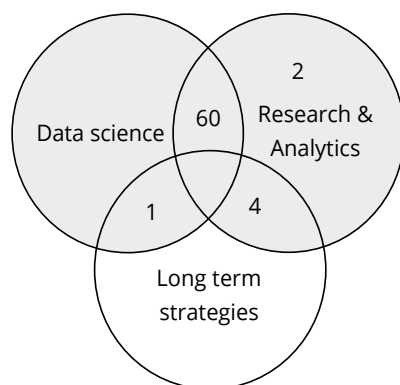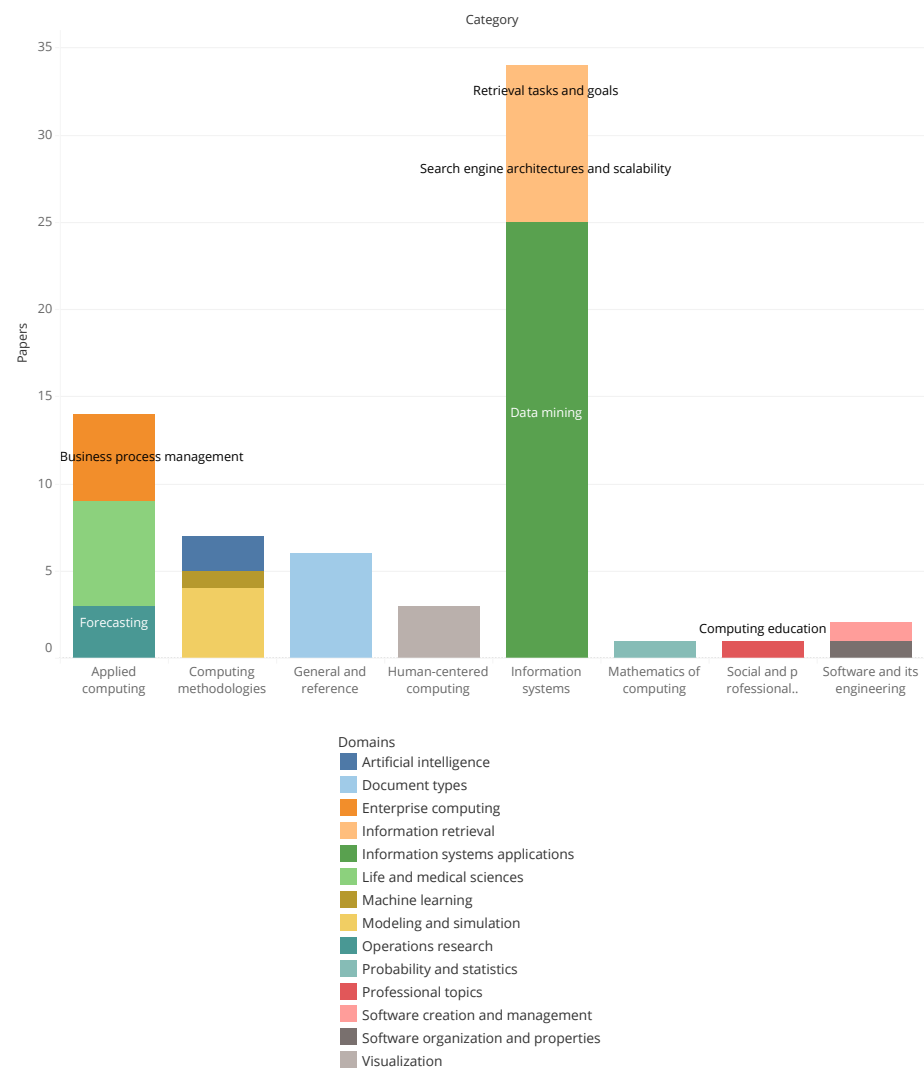| Taxonomy Description by Level	Elements
Applied computing	16
Enterprise computing	8
Business process management	7
Enterprise architectures	1
Life and medical sciences	6
Health care information systems	6
Operations research	3
Forecasting	2
Industry and manufacturing	1
Computing methodologies	7
Artificial intelligence	2
Natural language processing	2
Machine learning	1
Learning paradigms	1
Modeling and simulation	4
Model development and analysis	3
Simulation evaluation	1
General and reference	1
Cross-computing tools and techniques	1
Evaluation	1
Human-centered computing	3
Visualization	3
Visualization application domains	3
Information systems	36
Data management systems	1
Database design and models	1
Information retrieval	9
Retrieval tasks and goals	3
Search engine architectures and scalability	6
Information systems applications	26
Data mining	23
Decision support systems	3
Mathematics of computing	1
Probability and statistics	1
Statistical paradigms	1
Social and professional topics	1
Professional topics	1
Computing education	1
Software and its engineering	2
Software creation and management	1
Software verification and validation	1
Software organization and properties	1
Software system structures	1
Grand Total	67

Figure 2. Main topics by keywords.

In the trend analysis of the studies, we considered the year. The systematic review conducted in the SCOPUS database yielded results ranging from the year [2002, 2018]. Figure 4 shows the development and publication of all the content related to the topic, including those articles that we eliminated after applying the inclusion and exclusion criteria. We found the first publication in 2002, having a first peak during 2004 with twice as many publications as in the first year (2). We did not obtain any results for years in 2003 and 2006. The second peak occurred in 2007 with only five publications, doubling the number of publications compared to the previous peak. The third peak appears in 2009 after a decrease during 2008 to initial values (2); however, it starts to maintain what seems to be a sustained interest with an average of 5.4 publications between [2009, 2013]. In 2014 the first substantial increase was presented by going from 7 publications during 2013 to 13. This growth remains the following years; 2015 with 15 publications, 2016 with 17, 2017 and 2019 the years with the highest 39 and 18 publications. The range covers 16 years, from 2002 to 2018. The decline in 2018 may have been due to the query retrieval date, as the upward trend was evident. However, adding the most recent years [2019, 2021] could cause the trend to continue in an increasing function. This trend may suggest arisen interest in the research topic over time, as the difference concerning the first ten years 2002-2012 was still very minimal, from 2013 the number of publications had a drastic increase, with 2016 being the turning point where the number of publications doubled to the previous year. Figure 5 presents the articles included from two perspectives: document type and category. From this perspective, we can see that the Conference paper type of the Information systems type is the most common, followed by the Article type, which implies a journal publication and belongs to the same category. The article in press is present in two categories Human-centered computing and information systems; however, it is the least representative document type. Figure 6 shows a tree chart with the publications per year where we can see which papers were published each year.

Queries 1 and 2 in Table 1 give 37 results, in which 29 were included. From these results, none included or described the application of any strategy (either long- or short-term strategy). Due to the nature of the query, 23 of the 29 articles reviewed for this query involved some form of process mining technique, and one presented an event log aggregation technique. Table 5 shows the result of the techniques described in the articles. Discovery (15 articles) is the most common technique presented in the literature sample, followed by conformance checking (7 articles), which in most cases is described in conjunction with discovery.

Table 5. Results for query 1.

Process Mining Technique	Papers
Conformance	1
Discovery	15
Discovery, Conformance checking	7
Event log aggregation	1
No	5
Total	**29**

Figure 3. Taxonomy bar.

Figure 4. Number of publications per year.

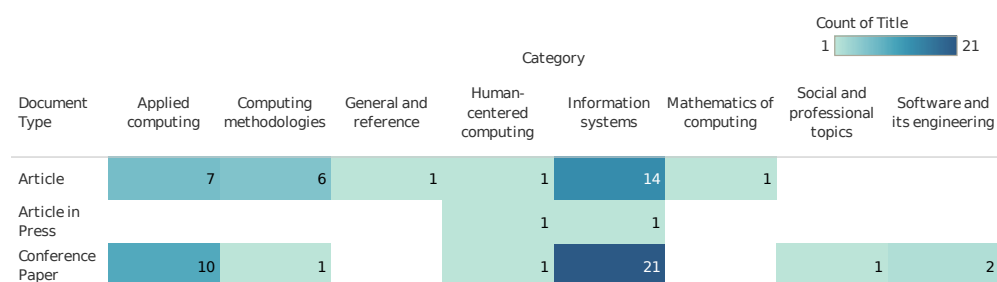| Document Type	Applied computing	Computing methodologies	General and reference	Human-centered computing	Information systems	Mathematics of computing	Social and professional topics	Software and its engineering
Article	7	6	1	1	14	1		
Article in Press				1	1			
Conference Paper	10	1		1	21		1	2

Figure 5. Number of publications per category.

4.3. Prediction Techniques

The objective of this observation is to verify how predictive techniques are used in conjunction with process mining. Dadashnia et al. [33] mentioned the possibility of predicting the next steps within an application using the conformance checking technique. However, they did not show the detail of how that activity would be performed. According to [15,34–37], a prediction is part of the background of the topic or literature. Flath and Stein [38] described that predictive analysis is a tool that should be taken into account in the manufacturing industry. They offered a set of tools and recommendations to facilitate this kind of machine learning application. Guidelines and best practices for modeling and interpreting features are also provided. However, the challenge remains an effective prediction problem within the particular context of the case study. Diapouli et al. [39] presented a case study with data mining algorithms for prediction tasks using a KNN algorithm.
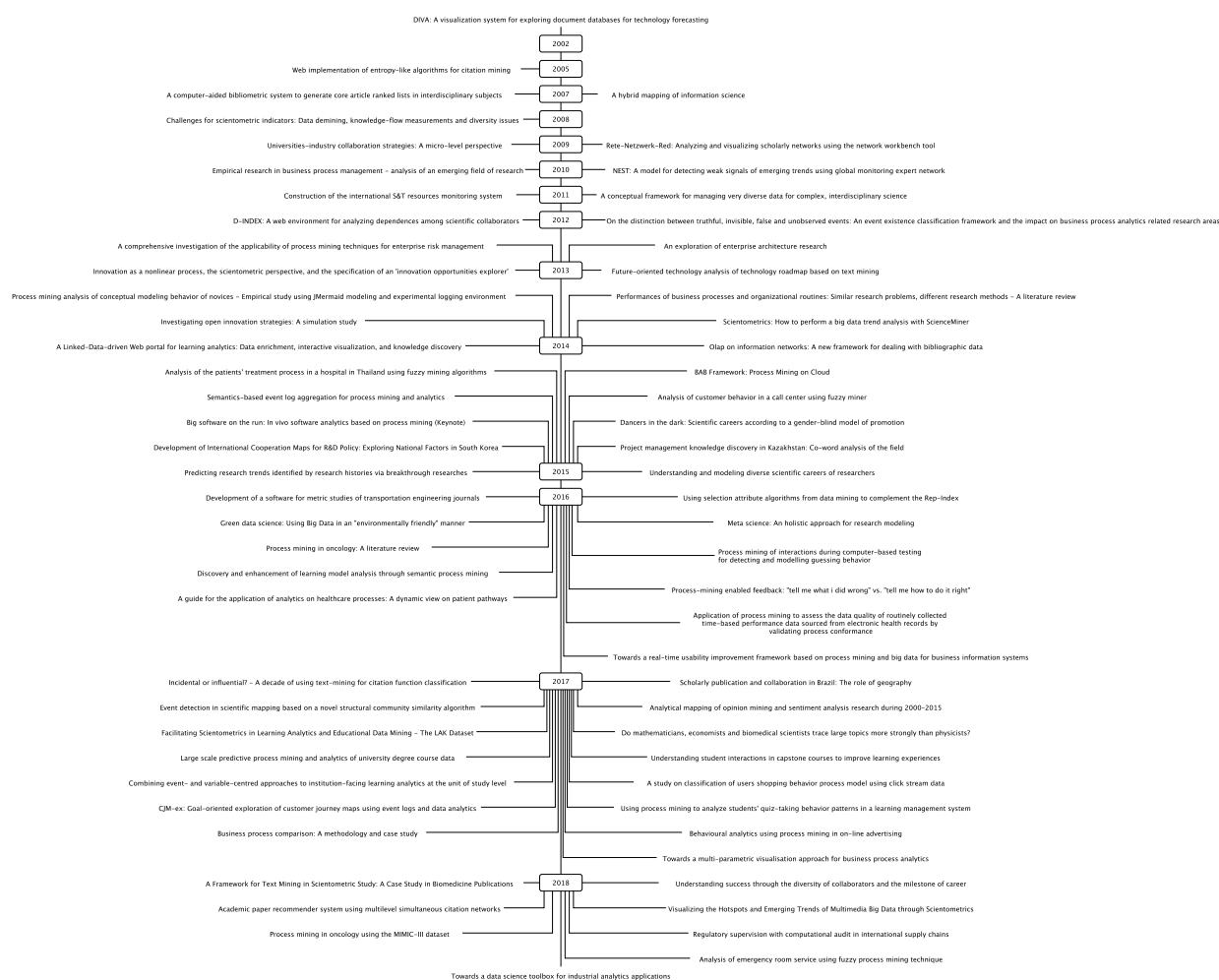
Figure 6. Tree chart with publications per year.

4.4. Recommendation

The following studies make some reference to recommendation tasks. Sedrakyan et al. [40] referred to the concept within the related study. Similarly, Neyem et al. [41] mentioned a future study to make recommendations using machine learning tools. Another study by Sedrakyan et al. [42] included empirical recommendations resulting from the experiments carried out. The recommendations were rather general from a mentoring perspective and recommendations for recording the student's data logs. Flath and Stein [38] mentioned only the association rules as part of the recommendation techniques within the unsupervised learning category. Son and Bum [43] propose collaborative and similar filtering methods for the recommendation of scientific papers. Their results are compared with those presented by tools such as Page Rank, which are superior in most cases.

4.5. Business Process

Sutrisnowati et al. [44] presented a web-based framework that implements process mining and distributed computing algorithms to deal with the information explosion in a business process analytics context. Syamsiyah et al. [45] introduced a new methodology to apply process comparison. Their methodology was successfully applied in a case study where a form dealing with a process was analyzed and actionable data were obtained by comparing different variants of the process using event data.

Caron et al. [46] provided a classification framework for events based on business criteria rather than technical aspects. The objective of this framework was to raise awareness about the different types of events that may exist in a business context. It removed the assumption that only events retrieved from event logs can be taken into account.

The framework provided a way to name schemes, resulting in more effective communication and better targeting of different business process analytics research areas. Caron et al. [47] presented an article that can serve as a reference for applying process mining to an industrial database. Wang et al. [48] described a framework that uses process mining, specifically discovery and conformance checking activities, to offer an auditing tool in supply chains to improve the administration of the process involved. The following study was retrieved about the context of the application. Bachhofner et al. [49] presented a multi-parametric visualization approach to address the lack of visualization principles applied to business process analytics. These principles were used to create, evaluate, and improve the approach in designing a process. They made a graphical representation of a process; however, they did not intend to emphasize process mining but rather described the area as a possible tool to use.

4.6. Domain Modeling

We retrieved two studies concerning domain modeling. The studies have the same main author. Sedrakyan et al. [42] presented an analysis of behavior modeling using an experimental logging function of the Jmermaid tool in conjunction with process mining techniques. The results included modeling patterns that were an indicator for better or worse learning performance. Guidelines for conceptual modeling focused on process-focused feedback were improved, and recommendations were provided for the type of data that may be useful in observing modeling behavior from the perspective of learning outcomes. Ideas were also provided for research analytics in the domain of conceptual modeling. Their next study [40] presented an approach for analyzing a learner's behavioral data in domain modeling. Discoveries included a set of modeling and validation patterns that can guide education for domain modeling courses. The study improved cognitive aspects of novice problem-solving behavior in domain modeling, specifically process-oriented feedback instead of traditional feedback.

4.7. Study of Behavior

There were five studies related to process mining in the study of behavior. Diapouli et al. [39] discussed applying process mining techniques and automatic learning based on online user behavior for predicting future behavior and biasing users by behavior. Once the behavior has been analyzed, the final task involved offering efficient advertising content. KNN and decision trees were used as prediction techniques. Padidem and Nalini [50] proposed the application of a business process management methodology. As input, there was an event log extracted from e-commerce sites. The objective was to classify the users using the purchase behavior using records of clicks flows and sequences of tasks. Four types of profiles and their real-time behavior were obtained using process mining techniques. Rattanathavorn and Premchaiswadi [51] performed an analysis of customer behavior within a call center. It was used as the main process mining technique, discovery, through the fuzzy miner algorithm. The objective of the behavior analysis involved knowing the dynamics of the process and the detection of characteristics that could be improved in favor of improving the service within that same context. Bernard and Andritsos [52] described customer journey mapping as an area of research related to user behavior when consuming a service. It provided a web interface that uses hierarchical clustering and statistical indices to enable interactive navigation through information stored in event logs. Juhaňák et al. [35] analyzed the behavior of students who perform activities based on quizzes using a learning management system (LMS). Kelly et al. [36] demonstrated the usefulness of combining event- and variable-based approaches when analyzing big data for higher education institutions. A dataset was used to demonstrate

the methodology. Results were presented about the relationships between student behavior in LMS and were explored using process modeling techniques.

4.8. Pattern Recognition

Papamitsiou and Economides [53] presented a process mining application for pattern recognition for the educational research community. Schulte et al. [54] performed a study to discover significant patterns in the study options a student will take, whose hypothesis was student-centered—the more informed you are about the effort it will take you to master a future subject, the better the study experience will be. They used process mining tools applied to educational research.

4.9. Software Engineering

There were the following studies related to process mining and software engineering. Neyem et al. [41] proposed a software tool to assist a course, providing support and ensuring compliance with most academic and engineering needs. The data extracted from this system were analyzed with process mining. The software was an integral solution for project management. Van der Aalst [37] analyzed the software "in vivo" to study systems in their natural habitat rather than through testing or software design. The objective was to observe systems running, obtain and analyze data from them, generate descriptive models, and use them to respond to failures. It focused on process mining as a tool for this activity, specifically discovery techniques.

4.10. Process Mining and Healthcare

The following study was determined within the context of healthcare. In [34], the authors discussed applying process mining in emergency rooms and improving procedures within this context. Jaroenphol et al. [55] performed a visualization to stimulate the patient's behavior extracted from an analyzed treatment process using a fuzzy miner algorithm. The results of the discovery process can be used as a tool to help researchers within the medical context to develop and improve hospital staff collaboration efficiently. Kurniati et al. [56] described the L*Lifecyle process and its application within the medical context. The case study corresponded to cancer pathways and attempted to demonstrate how process mining can be a useful tool using such a case study. Listmont et al. [57] presented a guide for the application of process analytics in healthcare. The aim was to use this guide for cases where you have processes related to healthcare and want to explore process mining. Perimal-Lewis et al. [58] described a process mining technology to assess data quality over time. Data were retrieved from an emergency department through electronic health records. Kurniati et al. [59] described a literary review on the application of mining processes to oncology cases.

4.11. Literature Review

The more extensive and complex of the two reviews corresponds to the paper presented by Houy et al. [60]. This article aims to contribute to the empirical research trends on Business Process Management (BPM). A systematic development framework attempts to identify the state of the art of the subject and its possible future development. To achieve this, it performs an exhaustive study of empirical research trends in BPM by analyzing formal contributions to science using scientometric methods. Through this study, the subject's maturity is presented by systematically analyzing a bibliographic selection with a well-defined and concise protocol, unlike the article presented by Breuker and Matzner [61], whose content is more related to a literature review than an exhaustive survey of the literature's methods. Moreover, this second article, rather than analyzing in-depth the available empirical research contributions and trends, aims to find a relationship between two topics addressed in the literature, on the one hand, Process Mining, as an area located within the general topic of BPM, and on the other hand the topic of organizational routines. From this

perspective, its purpose is to identify those works that could have some involvement in citation format between both topics.

One of the main differences is related to the data source. In contrast, in the first article, two bibliographic data sources are used—Science Citation Index (Thomson Scientific) and Business Source Premier (Ebsco). They use journal papers exclusively because the inclusion of conference proceedings could have reduced the representativeness of the literature review according to the year 2008 in which the Science Citation Index began to publish documents of this type, the second article uses Google Scholar and Scopus, without distinguishing between the two types of product. Finally, they reach different conclusions from the same common area—BPM. While the first article concludes with the assertion of the growing number of articles published on the subject, it underlines the increased interest in the field and the industry. The second article concludes with identifying a collaboration gap between Process Mining and organizational routines and suggests the possibility of adopting the methods and techniques of each area to each other, the result of which could generate innovative research with high potential.

4.12. Linked Data

There were two studies reviewed that were related to process mining and semantics. Deokar and Tao [62] proposed a framework for event log preprocessing focusing on event log aggregation. Phrase-based semantic similarity between normalized event names was used to add hierarchical event logs. Okoye et al. [63] described a semantic approach applied to process mining to enrich the event log streams of a learning process using semantic descriptions that reference concepts in an ontology specifically designed to represent the learning process. The approach involved the extraction of historical data from the process of learning execution environments. It demonstrated how data from a learning process could be semantically annotated and transformed into mine-capable event logs to predict individual patterns.

Hu et al. [64] presented a web portal based on linked data to apply scientometric methods. The web portal was based on a dataset called learning analytics and knowledge, which was already semantically annotated, so its structure was defined in a machine-readable way. Meanwhile, Dietze et al. [65] described a database of articles retrieved from query 5 related to learning analytics and educational data mining. The dataset has the particularity that was also annotated with semantic tags so that information was provided to facilitate the linking of data extracted from the metadata of the publications contained within the dataset. The objective was to offer a collection of publications mapped to an ontological vocabulary specifically for the learning analytics area.

4.13. Scientific Collaboration

Studies were selected for the topic of scientific collaboration. For example, Sidone et al. [66] provided information on collaborative networks and the impact that geographical location has on their constitution. Schifanella et al. [67] discussed the purpose of demonstrating the functionality of a web-based application that allows analyzing the degree of scientific collaboration between authors. With this tool, it is possible to identify the degree of importance that one author has had on another throughout his career and how much they have been mutually involved.

The study of Zitt and Bassecoulard [68] has as its main objective the description of the challenges facing the area of scientometrics, from the perspective of data "demining," knowledge-flow measurements, and diversity issues. For their part, Börner et al. offered a similar discussion about the challenges of studying science in a context that encompasses data mining, linguistics, and scientometrics. Another study that talks about future challenges within a scientific topic are found in Jin and Li [69], where they showed information related to the topic of multimedia big data to pave the way for the discovery of possible lines of research and gaps in the literature. Liu et al. [70] carried out a monitoring of the objectives of the global resources in science and technology, managing to identify the scientific

and technological frontiers through a systematic analysis of these resources. As a result, a set of solutions was defined, and techniques and methods were defined to be explored in more detail in the future. Li et al. [71] investigated the longitudinal tracing characteristics between scientists from different disciplines (physics, mathematics, biomedical sciences, and economics) to identify the differences between them concerning this characteristic. de Stefano et al. [72] presented a software tool that analyzes articles for the transportation engineering area.

4.14. Data Mining and Machine Learning

A significant part of the studies focused on using data mining techniques within a scientometric or similar context. For example, Xiangfeng [73] identified trends in research topics using clustering techniques; such trends were recognized through structural changes and various events observed within the dynamic behavior of clusters. Janssens et al. [74] also applied the clustering technique, but from a "hybrid" approach that integrated full text with citations and mapping of the field of information science. The hybrid method was compared against a traditional clustering and linear combination method to verify if the technique represented a better representation. Narbaev [75] described an exploratory article on the current state of research in project management in Kazakhstan; however, they used word analysis techniques and scientometric methods for the quantitative analysis of academic publications and cluster analysis of the most frequent publications related to project management. Morris et al. [76] presented software that performs bibliometric analysis whose results were presented by clusters in a two-dimensional visualization to explore the relationships between these clusters and the elements that compose them to support forecasting activities. Silalahi et al. [77] proposed a framework that allows text mining operations for scientometric studies using different classifiers such as Naive Bayes, k-NN, and SVM. Leydesdorff et al. [78] discussed how the innovation process followed nonlinear patterns in diverse science and technology. They described a multiperspective approach to reconstructing the stages of innovation applied to the RNA interference topic.

Cortés et al. [79] showed a web application of entropy algorithms aimed at text mining as part of the citation mining technique. Vivian et al. [80] used data mining techniques to propose a complement to the "Rep-Index" index, whose adaptations report better ratings compared with the original version. Ye and Feng [81] analyzed an algorithm to build a future-oriented technology and analyzed thesaurus of technology roadmap based on combining text mining and scientometrics with natural language. As a result of applying this algorithm, information was obtained about the development and characteristics of the technical fields to describe a technology analysis roadmap.

Cosentino et al. [82] described a conceptual scheme that provides a holistic view of information related to conferences to facilitating the study of manual retrieval of information from different resources. The scheme is capable of being used as a database for making queries and obtaining metrics. Guo [83] talked about a computer-aided bibliometric system to automatically generate a core article ranked list using scientometric indicators within different topics, such as mining and data and expert systems. Pride and Knoth [84] described a method of classification of scientific documents using the citations of a document through similarity metrics where one of the main results highlighted how the number of references in a document was a good indicator of prediction of the influence that the document will have.

4.15. Scientometrics and Trends

There were studies that mentioned trends and their combination with scientometrics. Yamashita et al. [85] proposed methods that allowed analysis of the scientific career and translated it into current research trends. Frehe et al. [86] presented an application to visualize trends in scientific topics. The aim was to offer a web tool that allows to search for contributions to a topic whose information was retrieved from different media, such as literature, news, or tweets. They also presented an analysis of the relationship between data

mining and scientometrics. Meanwhile, Jakawat et al. [87] described the current research trend but specifically for the topic of online analytical processing information networks. They presented a framework that builds several networks on the same topic, where each network constitutes a different perspective on how the topic was perceived or approached. Once the networks have been built, it was intended to store them in a data warehouse to apply tools for data mining, analysis, and visualization. Kim et al. [88] introduced the NEST model (new and emerging signals of trends) that collects information from experts around the world and detects weak trends but which allow new lines of research to be discovered through the foresight of future technology that supports decision-making about the implementation of R & D strategies. Pirayani et al. [89] described the details of a scientometric analysis of research on the subject of opinion mining and sentiment analysis.

4.16. Short and Long-Term Strategies

For example, Bjerregaard [90] discussed the importance of creating industry—university collaboration links. The main objective was to analyze collaboration strategies for optimizing the research and development (R and D) process. They also mentioned how short- and long-term strategies can help provide information about the collaborative relationship dynamics between institutions. However, the study was based on a qualitative study, which did not use data science techniques, nor did it offer a model of learning, recommendation, or prediction.

Fiegenbaum et al. [91] mentioned the innovation strategies defined in four profiles whose main characteristic was the variation of where it comes from and how knowledge was exploited. The profiles were defined in terms of open innovation (OI), closed innovation, outbound OI, and inbound OI. They made a model based on a simulation through a system based on agents. The objective of the simulation was to observe the competitive performance of each profile. The result of the observation provided the main one: the payoff obtained by certain profiles was not static but varied over time, that is, some of the strategies were more profitable in a short time than the others. Hence, the results involved finding out which strategy will be beneficial in the long-term and will be more profitable in the short-term. The results were not explicitly used as a prediction or recommendation mechanism, and using data science-related techniques was not described.

Although the study of Ramos et al. [92] mentioned the researcher's career, it rather described the differences between the scientific careers of men and women and how certain conditions, such as family members, affected their trajectories. The main result was the evidence of the linear and nonlinear difference of the trajectories for each gender; that is, while men have linear scientific careers, the pattern followed by women is far from it.

Yoon and Jeong [93] described the lack of long-term strategies for R & D collaboration between countries. They presented maps of cooperation between countries but considering that they were directly related to South Korea. The maps were obtained by analyzing international cooperation factors to create international strategies for groups of nations. They proposed establishing a customized strategy for each group of nations or countries determined by the collaborative map based on those factors that impact technology and the market.

Parsons et al. [94] presented strategies based on observation to facilitate socio-technical evolution in the entire data ecosystem. They presented a challenge in analyzing the diversity of interdisciplinary data, notably research data, its correlation and impact. It described the vision of data collections with the following characteristics: usability, simplicity, security, openness, and integrated connection to a framework that suggests a short- or long-term strategy of the data science ecosystem. It also described each of the characteristics that they considered the data should have as principles, challenges, and needs and compatibility of each characteristic. According to the observations presented, short-term strategies were suggested, which were related to the data systems and the entities that revolve around this topic. Similarly, long-term strategies were presented, which were in no way strategies extracted through data science techniques, but rather the observation result. For example,

they mentioned that data scientists need to continue with the professionalization of their discipline as a long-term recommendation.

The study of Bu et al. [95] is related to scientific collaboration by exploring diversity in this topic by analyzing the scientific career in computational sciences through two indicators: the diversity of the research topic and its impact (h-index). The study yielded a set of strategies resulting from the analysis of a database of scientific publications from ACM, including citation relationships. The exercise focused on exploring the impact of researchers and the diversity of topics of their collaborators, the relationship between the two concepts, and the identification of patterns of scientific collaboration after PhD graduation. As results were observed patterns that can inspire strategies, where among the discoveries was identified how, for example, the authors with more collaborators have a higher level of sociability that can help them succeed or, for example, it is suggested that high-impact authors tend to pursue collaborations on diverse topics.

Another related study is presented in [96]. They described the success of a researcher based on selecting the research area and research topic and in terms of change in research topics throughout a professional career. They analyzed the publication strategies of successful researchers and identified the features that define them. They also presented a methodology for the research and monitoring of the scientific career by exploring scientific communities based on characteristics, such as the type of conference a researcher can attend. To achieve the clustering, they defined the diversity of an author to group those authors with the same diversity characteristics. Two types of entropy are described for the definition of diversity: (i) flat entropy (it does not capture the order of information). Although such an order is required to characterize the change in the research fields adopted by an author in different periods, a zero flat entropy would indicate that an author did not change the field. The higher the value, the greater the diversity of research topics. (ii) Window entropy: This measurement takes a window of dimension k. For this window, the flat entropy was calculated, and the mean of all entropy was also calculated. With these two kinds of entropy, the diversity of the authors was analyzed. According to the data, a confusion matrix was constructed, indicating the correlation between both types of entropy. The matrix corresponds to the low and high values for each type of entropy. Each cell in the matrix corresponds to a career profile. A low window entropy and high plain entropy describes authors who do not study simultaneously in multiple fields but one after another in a time slice. They proposed investigating the preferred strategy that a new author should adopt to increase relevance in the scientific community. To validate the approach, they implemented a stochastic model to reproduce an author's field selection and then evaluate the model's predictions against actual results. The best strategy detected was to study in multiple fields of research in the entire career, but to remain confined in a few fields in each time window with the advantage of having more relevance and getting a greater number of cites; however, this was the least common strategy. The worst strategy was represented by the profile (high plain entropy + high window entropy), indicating that it is working in many research fields simultaneously within each time window, making it one of the most popular strategies, but not the best. The approach presented in this study described the following application proposals—labeling of publications, analyzing professional growth by observing the adaptation process, designing the best selection pattern for career development, and developing collaborative forecasting systems that can recommend a particular partner to collaboration.

5. Discussion

5.1. Summary of Evidence

We have described the recommendation techniques regarding data mining in the literature. Within the sample, studies were found that referred to this concept based on using machine learning algorithms. However, the detected recommendations are usually empirical and more general. The recommendations are based on the data analyzed in

the articles but are rather the result of a hybrid approach (i.e., observation of data and suggestions for recommendations).

Based on the results analyzed, no approaches using process mining techniques as learning tools for long-term strategies have been observed. Most of the studies involved applying process mining techniques where the most common is represented by discovery. However, in the literature, there are cases of prediction techniques using mining processes through techniques, such as conformance checking, which opens the possibility to the application of an objective function that can be predicted for the next event. This opens up the possibilities of studying the effect of executing one action on another or that an event occurs on another and its influence on a measurable result.

Although the found prediction techniques emphasize machine learning algorithms, such as decision trees and KNN, mainly for behavior prediction, it is possible to use process mining algorithms with a similar objective. This type of approach is designed to analyze and model behavioral patterns based on them, offering content.

Using process mining in varied contexts has been detected in the sample. The application of process mining covers topics from healthcare to software engineering. However, no study is detected that is explicitly related to research analytics; the closest concept is related to educational research. Although studies related to scientific collaboration are found, their most explicit relationship can be found directly with data mining, such as identifying trends in research topics, clustering techniques for citation information, and quantitative analyses of scientific output.

Although there are studies that mention long- and short-term strategies, they are mostly strategies that were analyzed on the basis of qualitative studies, not using data mining or process mining techniques to obtain them, but rather are part of an empirical evaluation of the results obtained. This implies that although the strategies are part of a data exploration methodology, the results are not explicitly obtained by creating automatic recommendation mechanisms (i.e., the strategies are inspired by identifying observed patterns). However, a related case was found where an entropy technique is used to define production patterns and, based on them, create a set of user profiles to define strategies that lead to improving the researcher's career. Although it corresponds to one of the most related studies or those expected to be found to a greater extent, it does not use process mining techniques. However, it stands out for its use to develop forecasting system techniques that can help make collaborative recommendations.

5.2. Limitations

There are limitations in this study. As previously described, the systematic review is based on the methodologies described in [20,31]. Some of the limitations detected are described below. First, the sample was obtained from a single bibliometric database (Scopus). Although the articles come from multiple publishers, the document retrieval queries were only consulted in that source. Although only libraries accessible through the university network were used (e.g., Science Direct, Springer Link, and IEEE Xplore), those studies that were not accessible through the university network were requested directly from the authors. Second, to create the data retrieval queries, we defined a set of keywords, with which a combination was made to bring the greatest number of results using these words. However, only those queries that returned results were presented in this study. All those queries that returned 0 or no results were omitted; currently, there are 48 queries made using the keywords. Third, the review and selection of articles were broad to group a larger number of articles directly related to the keywords used. Four, another limitation related to SCOPUS is the prevalence of the English language and the elimination of papers in the researchers' local language. Petrushka et al. [97] refer to this issue, where they mention that English content occupies a dominant position in the scientific literature, displacing content in other languages.

6. Conclusions

The objectives of this systematic review are (i) to identify whether methods existed for recommending long-term strategies using data science techniques, where one of the main techniques had to be directly related to process mining, and (ii) to identify whether there were items that used process mining within the context of research analytics. The search space that integrates a bibliometric database (Scopus) resulted in 67 selected articles, whose articles were very varied, to identify the previously established. Based on the articles reviewed, only a small sample of them mention recommendation strategies in the short and long terms; however, they do not use process mining techniques and are based on the observation of results. Only one of them is based on the application of data mining within its methodology, so it is possible to create production profiles that can be used as strategies in the short and long terms. Our future study is heading in the following directions. This leaves open research opportunities from different perspectives: applying methodologies involving process mining for the context of research analytics and the feasibility study on long-term strategies using process mining techniques. Results suggest having open research opportunities from different perspectives: applying methodologies involving process mining for the context of research analytics and the feasibility study on long-term strategies using process mining techniques and including results of articles published up to 2021.

Abbreviations

ACM	Association for Computing Machinery
BPM	Business Process Management
BPMN	Business Process Model and Notation
CONACYT	National Council of Science and Technology
DOAJ	Directory of open access journals
MDPI	Multidisciplinary Digital Publishing Institute
LMS	Learning management system

References

1. Sánchez-Torres, J.M.; Miles, I. The role of future-oriented technology analysis in e-Government: A systematic review. *Eur. J. Futur. Res.* **2017**, *5*, 15. [CrossRef]
2. Galey, S.A.; Lerner, Z.F.; Bulea, T.C.; Zimbler, S.; Damiano, D.L. Effectiveness of surgical and non-surgical management of crouch gait in cerebral palsy: A systematic review. *Gait Posture* **2017**, *54*, 93–105. [CrossRef] [PubMed]
3. Georgia, M.; Kaye, M.; Darlene, M.; Sue, B. The experiences and perceptions of food banks amongst users in high-income countries: An international scoping review. *Appetite* **2018**, *120*, 698–708. [CrossRef]

4. Herrera, S.; Enuameh, Y.; Adjei, G.; Ae-Ngibise, K.A.; Asante, K.P.; Sankoh, O.; Owusu-Agyei, S.; Yé, Y. A systematic review and synthesis of the strengths and limitations of measuring malaria mortality through verbal autopsy. *Malar. J.* **2017**, *16*, 421. [CrossRef] [PubMed]

5. Simpkin, A.L.; Robertson, L.C.; Barber, V.S.; Young, J.D. Modifiable factors influencing relatives' decision to offer organ donation: Systematic review. *BMJ* **2009**, *338*. [CrossRef] [PubMed]

6. Ryde, G.C.; Gilson, N.D.; Burton, N.W.; Brown, W.J. Recruitment Rates in Workplace Physical Activity Interventions: Characteristics for Success. *Am. J. Health Promot.* **2013**, *27*, e101–e112. [CrossRef]

7. Haines, M.L.; Anderson, R.P.; Gibson, P.R. Systematic review: The evidence base for long-term management of coeliac disease. *Aliment. Pharmacol. Ther.* **2008**, *28*, 1042–1066. [CrossRef]

8. Gómez, G.; Aguileta, A.; Ancona, G.B.; Gómez, O.S. Avances en las Mejoras de Procesos Software en las MiPyMEs Desarrolladoras de Software: Una Revisión Sistemática. *Rev. Latinoam. Ing. Softw.* **2014**, *2*, 262–268. [CrossRef]

9. Leydesdorff, L.; Milojević, S. Scientometrics, In *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed.; Elsevier: Oxford, UK, 2015. [CrossRef]

10. Cantu-Ortiz, F.J. *Research Analytics: Boosting University Productivity and Competitiveness through Scientometrics*; Auerbach Publications: New York, NY, USA, 2018; pp. 1–264.

11. Najafabadi, M.; Villanustre, F.; Khoshgoftaar, T.; Seliya, N.; Wald, R.; Muharemagic, E. Deep learning applications and challenges in big data analytics. *J. Big Data* **2015**, *2*. [CrossRef]

12. Van der Aalst, W. *Process Mining: Data Science in Action*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2016.

13. Waller, M.; Fawcett, S. Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *J. Bus. Logist.* **2013**, *34*, 77–84. [CrossRef]

14. Provost, F.; Fawcett, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* **2013**, *1*, 51–59. [CrossRef] [PubMed]

15. Van Der Aalst, W. Green data science: Using Big Data in an "environmentally friendly" manner. In Proceedings of the ICEIS 2016—Proceedings of the 18th International Conference on Enterprise Information Systems, Rome, Italy, 25–28 April 2016; Volume 1, pp. 9–21.

16. Van der Aalst, W.M.P. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1 ed.; Springer: Berlin, Germany, 2011; p. 352. [CrossRef]

17. Dijkman, R.M.; Dumas, M.; Ouyang, C. Semantics and analysis of business process models in BPMN. *Inf. Softw. Technol.* **2008**, *50*, 1281–1294. [CrossRef]

18. Mitchell, T.M. *Machine Learning*, 1 ed.; McGraw-Hill, Inc.: New York, NY, USA, 1997.

19. Alpaydin, E. *Introduction to Machine Learning*, 2nd ed.; The MIT Press: Cambridge, MA, USA, 2010.

20. De Almeida Biolchini, J.C.; Mian, P.G.; Natali, A.C.C.; Travassos, G.H. *Systematic Review in Software Engineering*; Technical Report; Alberto Luiz Coimbra Institute for Graduate Studies and Research in Engineerin COPPE/UFJR: Rio de Janeiro, Brazil, 2005.

21. Murad, M.H.; Montori, V.M.; Ioannidis, J.P.A.; Jaeschke, R.; Devereaux, P.J.; Prasad, K.; Neumann, I.; Carrasco-Labra, A.; Agoritsas, T.; Hatala, R.; et al. How to Read a Systematic Review and Meta-analysis and Apply the Results to Patient Care: Users' Guides to the Medical Literature. *JAMA* **2014**, *312*, 171–179. [CrossRef]

22. Korevaar, D.A.; Hooft, L.; Riet, G.T. Systematic reviews and meta-analyses of preclinical studies: Publication bias in laboratory animal experiments. *Lab. Anim.* **2011**, *45*, 225–230. [CrossRef]

23. Peters, J.L.; Sutton, A.J.; Jones, D.R.; Rushton, L.; Abrams, K.R. A Systematic Review of Systematic Reviews and Meta-Analyses of Animal Experiments with Guidelines for Reporting. *J. Environ. Sci. Health Part B* **2006**, *41*, 1245–1258. [CrossRef]

24. Meerpohl, J.J.; Herrle, F.; Antes, G.; von Elm, E. Scientific Value of Systematic Reviews: Survey of Editors of Core Clinical Journals. *PLoS ONE* **2012**, *7*, e35732. [CrossRef]

25. Petticrew, M. Systematic reviews from astronomy to zoology: Myths and misconceptions. *BMJ* **2001**, *322*, 98–101. [CrossRef] [PubMed]

26. Mulrow, C.D. Systematic Reviews: Rationale for systematic reviews. *BMJ* **1994**, *309*, 597–599. [CrossRef] [PubMed]

27. Knipschild, P. Systematic Reviews: Some examples. *BMJ* **1994**, *309*, 719–721. [CrossRef] [PubMed]

28. Robinson, P.; Lowe, J. Literature reviews vs systematic reviews. *Aust. N. Z. J. Public Health* **2015**, *39*, 103. [CrossRef] [PubMed]

29. Egger, M.; Smith, G.D.; Phillips, A.N. Meta-analysis: Principles and procedures. *BMJ* **1997**, *315*, 1533–1537. [CrossRef] [PubMed]

30. Egger, M.; Smith, G.D. Meta-analysis: Potentials and promise. *BMJ* **1997**, *315*, 1371–1374. [CrossRef] [PubMed]

31. Petticrew, M.; Roberts, H. *Systematic Reviews in the Social Sciences: A Practical Guide*; Blackwell Publishing Ltd.: Hoboken, NJ, USA, 2008; pp. 1–336. [CrossRef]

32. The 2012 ACM Computing Classification System Home Page. Association for Computing Machinery. Available online: https://www.acm.org/publications/class-2012 (accessed on 18 September 2018).

33. Dadashnia, S.; Niesen, T.; Fettke, P.; Loos, P. Towards a Real-time Usability Improvement Framework based on Process Mining and Big Data for Business Information Systems. In Proceedings of the Tagungsband Multikonferenz Wirtschaftsinformatik (MKWI-16), Ilmenau, Germany, 9–11 March 2016.

34. Jangvaha, K.; Porouhan, P.; Palangsantikul, P.; Premchaiswadi, W. Analysis of emergency room service using fuzzy process mining technique. In Proceedings of the 2017 15th International Conference on ICT and Knowledge Engineering (ICT KE), Bangkok, Thailand, 24–26 November 2017; pp. 1–5. [CrossRef]

35. Juhaňák, L.; Zounek, J.; Rohlíková, L. Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Comput. Hum. Behav.* **2017**. [CrossRef]

36. Kelly, N.; Montenegro, M.; Gonzalez, C.; Clasing, P.; Sandoval, A.; Jara, M.; Saurina, E.; Alarcón, R. Combining event- and variable-centred approaches to institution-facing learning analytics at the unit of study level. *Int. J. Inf. Learn. Technol.* **2017**, *34*, 63–78. [CrossRef]

37. Aalst, W.v.d. Big Software on the Run: In Vivo Software Analytics Based on Process Mining (Keynote). In Proceedings of the 2015 International Conference on Software and System Process, Tallinn, Estonia, 24–26 August 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 1–5. [CrossRef]

38. Flath, C.M.; Stein, N. Towards a data science toolbox for industrial analytics applications. *Comput. Ind.* **2018**, *94*, 16–25. [CrossRef]

39. Diapouli, M.; Kapetanakis, S.; Petridis, M.; Evans, R. Behavioural analytics using process mining in on-line advertising. In Proceedings of the ICCBR 2017 Workshops, Trondheim, Norway, 26–28 June 2017; Volume 2028, pp. 147–156.

40. Sedrakyan, G.; De Weerdt, J.; Snoeck, M. Process-mining enabled feedback: "Tell me what i did wrong" vs. "tell me how to do it right". *Comput. Hum. Behav.* **2016**, *57*, 352–376. [CrossRef]

41. Neyem, A.; Diaz-Mosquera, J.; Munoz-Gama, J.; Navon, J. Understanding Student Interactions in Capstone Courses to Improve Learning Experiences. In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, Seattle, WA, USA, 8–11 March 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 423–428. [CrossRef]

42. Sedrakyan, G.; Snoeck, M.; De Weerdt, J. Process mining analysis of conceptual modeling behavior of novices—Empirical study using JMermaid modeling and experimental logging environment. *Comput. Hum. Behav.* **2014**, *41*, 486–503. [CrossRef]

43. Son, J.; Kim, S. Academic paper recommender system using multilevel simultaneous citation networks. *Decis. Support Syst.* **2018**, *105*, 24–33. [CrossRef]

44. Sutrisnowati, R.A.; Bae, H.; Pulshashi, I.R.; Putra, A.D.A.; Prastyabudi, W.A.; Park, S.; Joo, B.; Choi, Y. BAB Framework: Process Mining on Cloud. *Procedia Comput. Sci.* **2015**, *72*, 453–460. [CrossRef]

45. Syamsiyah, A.; Bolt, A.; Cheng, L.; Hompes, B.F.A.; Jagadeesh Chandra Bose, R.P.; van Dongen, B.F.; van der Aalst, W.M.P. Business Process Comparison: A Methodology and Case Study. In *Business Information Systems*; Abramowicz, W., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 253–267.

46. Caron, F.; vanden Broucke, S.; Vanthienen, J.; Baesens, B. On the distinction between truthful, invisible, false and unobserved events: An event existence classification framework and the impact on business process analytics related research areas. In Proceedings of the AMCIS 2012, Seattle, WA, USA, 9–11 August 2012; Volume 1, pp. 50–60.

47. Caron, F.; Vanthienen, J.; Baesens, B. A comprehensive investigation of the applicability of process mining techniques for enterprise risk management. *Comput. Ind.* **2013**, *64*, 464–475. [CrossRef]

48. Wang, Y.; Hulstijn, J.; Tan, Y.h. Regulatory Supervision with Computational Audit in International Supply Chains. In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, Delft, The Netherlands, 30 May 2018; Association for Computing Machinery: New York, NY, USA, 2018, [CrossRef]

49. Bachhofner, S.; Kis, I.; Di Ciccio, C.; Mendling, J. Towards a Multi-parametric Visualisation Approach for Business Process Analytics. In Proceedings of the Advanced Information Systems Engineering Workshops, Essen, Germany, 12–16 June 2017; pp. 85–91.

50. Padidem, D.; Nalini, C. A study on classification of users shopping behavior process model using click stream data. *J. Eng. Appl. Sci.* **2017**, *12*, 9548–9553. [CrossRef]

51. Rattanathavorn, K.; Premchaiswadi, W. Analysis of customer behavior in a call center using fuzzy miner. In Proceedings of the 2015 13th International Conference on ICT and Knowledge Engineering (ICT Knowledge Engineering 2015), Bangkok, Thailand, 18–20 November 2015; pp. 137–141. [CrossRef]

52. Bernard, G.; Andritsos, P. CJM-ex: Goal-oriented Exploration of Customer Journey Maps using Event Logs and Data Analytics. In Proceedings of the BPM Demo Track and BPM Dissertation Award Co-Located with 15th International Conference on Business Process Modeling (BPM 2017), Barcelona, Spain, 10–15 September 2017; Volume 1920.

53. Papamitsiou, Z.; Economides, A. Process mining of interactions during computer-based testing for detecting and modelling guessing behavior. *Lect. Notes Comput. Sci.* **2016**, *9753*, 437–449. [CrossRef]

54. Schulte, J.; Fernandez de Mendonca, P.; Martinez-Maldonado, R.; Buckingham Shum, S. Large Scale Predictive Process Mining and Analytics of University Degree Course Data. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, BC, Canada, 13–17 March 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 538–539. [CrossRef]

55. Jaroenphol, E.; Porouhan, P.; Premchaiswadi, W. Analysis of the patients' treatment process in a hospital in Thailand using fuzzy mining algorithms. In Proceedings of the 2015 13th International Conference on ICT and Knowledge Engineering (ICT Knowledge Engineering 2015), Bangkok, Thailand, 18–20 November 2015; pp. 131–136. [CrossRef]

56. Kurniati, A.P.; Geoff, H.; David, H.; Owen, J. Process mining in oncology using the MIMIC-III dataset. *J. Phys. Conf. Ser.* **2018**, *971*, 012008. [CrossRef]

57. Lismont, J.; Janssens, A.S.; Odnoletkova, I.; vanden Broucke, S.; Caron, F.; Vanthienen, J. A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways. *Comput. Biol. Med.* **2016**, *77*, 125–134. [CrossRef]

58. Perimal-Lewis, L.; Teubner, D.; Hakendorf, P.; Horwood, C. Application of process mining to assess the data quality of routinely collected time-based performance data sourced from electronic health records by validating process conformance. *Health Inform. J.* **2016**, *22*, 1017–1029. [CrossRef]

59. Kurniati, A.P.; Johnson, O.; Hogg, D.; Hall, G. Process mining in oncology: A literature review. In Proceedings of the 2016 6th International Conference on Information Communication and Management (ICICM), Hertfordshire, UK, 29–31 October 2016; IEEE: Hatfield, UK, 2016; pp. 291–297. [CrossRef]

60. Houy, C.; Fettke, P.; Loos, P. Empirical research in business process management - analysis of an emerging field of research. *Bus. Process. Manag. J.* **2010**, *16*, 619–661. [CrossRef]

61. Breuker, D.; Matzner, M. Performances of business processes and organizational routines: Similar research problems, different research methods—A literature review. In Proceedings of the 22nd European Conference on Information Systems, Tel Aviv, Israel, 9–11 June 2014.

62. Deokar, A.V.; Tao, J. Semantics-based event log aggregation for process mining and analytics. *Inf. Syst. Front.* **2015**, *17*, 1209–1226. [CrossRef]

63. Okoye, K.; Tawil, A.R.; Naeem, U.; Lamine, E. Discovery and enhancement of learning model analysis through semantic process mining. *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.* **2016**, *8*, 93–114.

64. Hu, Y.; McKenzie, G.; Yang, J.A.; Gao, S.; Abdalla, A.; Janowicz, K. A Linked-Data-driven Web portal for learning analytics: Data enrichment, interactive visualization, and knowledge discovery. In Proceedings of the CEUR Workshop Proceedings, Indianapolis, IN, USA, 24–28 March, 2014; Volume 1137.

65. Dietze, S.; Taibi, D.; D'Aquin, M. Facilitating Scientometrics in Learning Analytics and Educational Data Mining—The LAK Dataset. *Semant. Web* **2017**, *8*, 395–403. [CrossRef]

66. Sidone, O.; Haddad, E.; Mena-Chalco, J. Scholarly publication and collaboration in Brazil: The role of geography. *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 243–258. [CrossRef]

67. Schifanella, C.; Di Caro, L.; Cataldi, M.; Aufaure, M.A. D-INDEX: A Web Environment for Analyzing Dependences Among Scientific Collaborators. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; ACM: New York, NY, USA, 2012; pp. 1520–1523. [CrossRef]

68. Zitt, M.; Bassecoulard, E. Challenges for scientometric indicators: Data demining, knowledge-flow measurements and diversity issues. *Ethics Sci. Environ. Politics* **2008**, *8*, 49–60. [CrossRef]

69. Jin, Y.; Li, X. Visualizing the Hotspots and Emerging Trends of Multimedia Big Data through Scientometrics. *Multimed. Tools Appl.* **2018**, 1–25. [CrossRef]

70. Liu, Y.; Wang, X.L.; Wang, W.P.; Ye, X.T.; Fan, W. Construction of the international S&T resources monitoring system. *Commun. Comput. Inf. Sci.* **2011**, *238 CCIS*, 343–355. [CrossRef]

71. Li, M.; Yang, L.; Zhang, H.; Shen, Z.; Wu, C.; Wu, J. Do mathematicians, economists and biomedical scientists trace large topics more strongly than physicists? *J. Inf.* **2017**, *11*, 598–607. [CrossRef]

72. de Stefano, E.; de Sequeira Santos, M.P.; Balassiano, R. Development of a software for metric studies of transportation engineering journals. *Scientometrics* **2016**, *109*, 1579–1591. [CrossRef]

73. Xiangfeng, M.; Xinhai, L.; Yan, Z.; Wolfgang, G. Event detection in scientific mapping based on a novel structural community similarity algorithm. In Proceedings of the 16th International Conference on Scientometrics and Informetrics, Wuhan,China, 16–20 October 2017; pp. 258–269.

74. Janssens, F.; Glänzel, W.; De Moor, B. A hybrid mapping of information science. *Scientometrics* **2008**, *75*, 607–631. [CrossRef]

75. Narbaev, T. Project management knowledge discovery in Kazakhstan: Co-word analysis of the field. In Proceedings of the 12th International Conference on Intellectual Capital, Knowledge Management & Organisational Learning, Bangkok, Thailand, 5–6 November 2015; Volume 2015, pp. 169–175.

76. Morris, S.; DeYong, C.; Wu, Z.; Salman, S.; Yemenu, D. DIVA: A visualization system for exploring document databases for technology forecasting. *Comput. Ind. Eng.* **2002**, *43*, 841–862. [CrossRef]

77. Silalahi, V.M.M.; Hardiyati, R.; Nadhiroh, I.M.; Handayani, T.; Rahmaida, R.; Amelia, M. A Framework for Text Mining in Scientometric Study: A Case Study in Biomedicine Publications. *J. Phys. Conf. Ser.* **2018**, *1007*, 012030. [CrossRef]

78. Leydesdorff, L.; Rotolo, D.; de Nooy, W. Innovation as a nonlinear process, the scientometric perspective, and the specification of an 'innovation opportunities explorer'. *Technol. Anal. Strateg. Manag.* **2013**, *25*, 641–653. [CrossRef]

79. Cortés, H.; Del Río, J.; García, E. Web implementation of entropy-like algorithms for citation mining. *WSEAS Trans. Inf. Sci. Appl.* **2005**, *2*, 1430–1437.

80. Vivian, G.; Cervi, C.; Rovadosky, D. Using selection attribute algorithms from data mining to complement the Rep-Index. In Proceedings of the 15th International Conference WWW/Internet 2016, Mannheim, Germany, 28–30 October 2016; pp. 219–226.

81. Ye, C.; Feng, L. Future-oriented technology analysis of technology roadmap based on text mining. In Proceedings of the 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Shenyang, China, 23–25 July 2013; pp. 1126–1130. [CrossRef]

82. Cosentino, V.; Izquierdo, J.; Cabot, J. Meta science: An holistic approach for research modeling. *Lect. Notes Comput. Sci.* **2016**, *9974 LNCS*, 365–380. [CrossRef]

83. Guo, G. A computer-aided bibliometric system to generate core article ranked lists in interdisciplinary subjects. *Inf. Sci.* **2007**, *177*, 3539–3556. [CrossRef]

84. Pride, D.; Knoth, P. Incidental or influential? A decade of using text-mining for citation function classification. In Proceedings of the International Society of Scientometrics and Informetrics Conference 2017, Wuhan, China, 16–20 October 2017; pp. 1357–1367.

85. Yamashita, N.; Numao, M.; Ichise, R. Predicting research trends identified by research histories via breakthrough researches. *IEICE Trans. Inf. Syst.* **2015**, *E98D*, 355–362. [CrossRef]

86. Frehe, V.; Rugaitis, V.; Teuteberg, F. Scientometrics: How to perform a big data trend analysis with ScienceMiner. *Informatik* **2014**, *P-232*, 1699–1710.

87. Jakawat, W.; Favre, C.; Loudcher, S. Olap on information networks: A new framework for dealing with bibliographic data. *Adv. Intell. Syst. Comput.* **2014**, *241*, 361–370. [CrossRef]

88. Kim, S.; Kwon, Y.; Jeong, Y.; Choi, S.B.; Park, J.K.; Hong, S.W. NEST: A model for detecting weak signals of emerging trends using global monitoring expert network. In Proceedings of the EKAW, Lisbon, Portugal, 11–15 October 2010; Volume 674.

89. Piryani, R.; Madhavi, D.; Singh, V. Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Inf. Process. Manag.* **2017**, *53*, 122–150. [CrossRef]

90. Bjerregaard, T. Universities-industry collaboration strategies: A micro-level perspective. *Eur. J. Innov. Manag.* **2009**, *12*, 161–176. [CrossRef]

91. Fiegenbaum, I.; Ihrig, M.; Torkkeli, M. Investigating open innovation strategies: A simulation study. *Int. J. Technol. Manag.* **2014**, *66*, 183–211. [CrossRef]

92. González Ramos, A.M.; Navarrete Cortés, J.; Cabrera Moreno, E. Dancers in the Dark: Scientific Careers According to a Gender-Blind Model of Promotion. *Interdiscip. Sci. Rev.* **2015**, *40*, 182–203. [CrossRef]

93. Yoon, B.; Jeong, Y. Development of International Cooperation Maps for R&D Policy: Exploring National Factors in South Korea. *Sci. Technol. Soc.* **2015**, *20*, 225–251. [CrossRef]

94. Parsons, M.A.; Godøy, Ø.; LeDrew, E.; de Bruin, T.F.; Danis, B.; Tomlinson, S.; Carlson, D. A conceptual framework for managing very diverse data for complex, interdisciplinary science. *J. Inf. Sci.* **2011**, *37*, 555–569. [CrossRef]

95. Bu, Y.; Ding, Y.; Xu, J.; Liang, X.; Gao, G.; Zhao, Y. Understanding success through the diversity of collaborators and the milestone of career. *J. Assoc. Inf. Sci. Technol.* **2018**, *69*, 87–97. [CrossRef]

96. Chakraborty, T.; Tammana, V.; Ganguly, N.; Mukherjee, A. Understanding and modeling diverse scientific careers of researchers. *J. Inf.* **2015**, *9*, 69–78. [CrossRef]

97. Petrushka, A.; Komova, M.; Fedushko, S. Scientific Content: Language Expansion in Bibliometric Databases. In Proceedings of the International Workshop on Cyber Hygiene (CybHyg-2019), Lviv, Ukraine, 30 November 2020; Volume 2654, pp. 375–389.