

Review Article

Recommendation and Classification Systems: A Systematic Mapping Study

J. G. Enríquez ¹, **L. Morales-Trujillo** ¹, **Fernando Calle-Alonso** ²,
F. J. Domínguez-Mayo ¹ and **J. M. Lucas-Rodríguez**³

¹Computer Languages and Systems Department, University of Seville, Avd. Reina Mercedes s/n, 41012 Seville, Spain

²Statistics and Operational Research Area, University of Malaga, Bulevar Louis Pasteur 31, 29010 Malaga, Spain

³Servinform S.A., Calle Manufactura, 5, 41927 Mairena del Aljarafe, Spain

Correspondence should be addressed to J. G. Enríquez; jose.gonzalez@iwt2.org

Received 15 February 2019; Revised 14 May 2019; Accepted 12 June 2019; Published 27 June 2019

Guest Editor: Luis Iribarne

Copyright © 2019 J. G. Enríquez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Today, recommendation algorithms are widely used by companies in multiple sectors with the aim of increasing their profits or offering a more specialized service to their customers. Moreover, there are countless applications in which classification algorithms are used, seeking to find patterns that are difficult for people to detect or whose detection cost is very high. Sometimes, it is necessary to use a mixture of both algorithms to give an optimal solution to a problem. This is the case of the ADAGIO, a R&D project that combines machine learning (ML) strategies from heterogeneous data sources to generate valuable knowledge based on the available open data. In order to support the ADAGIO project requirements, the main objective of this paper is to provide a clear vision of the existing classification and recommendation ML systems to help researchers and practitioners to choose the best option. To achieve this goal, this work presents a systematic review applied in two contexts: scientific and industrial. More than a thousand papers have been analyzed resulting in 80 primary studies. Conclusions show that the combination of these two algorithms (classification and recommendation) is not very used in practice. In fact, the validation presented for both cases is very scarce in the industrial environment. From the point of view of software development life cycle, this review also shows that the work being done in the ML (for classification and recommendation) research and industrial environment is far from earlier stages such as business requirements and analysis. This makes it very difficult to find efficient and effective solutions that support real business needs from an early stage. It is therefore that the article suggests the development of new ML research lines to facilitate its application in the different domains.

1. Introduction

The great growth in the amount of data and information that can be accessed (the known Big Data), coupled with government collaboration to provide open information (Open Data), makes companies very interested in this issue. One of the biggest problems in this area is that this information is not found in one single place, not even in a common interpretation format. Therefore, it is necessary to create solutions that collect these dispersed data and apply a specific treatment so that they can be offered to their customers.

The collection of dispersed information and its unification in order to be able to work with it would open a new market niche, a new business unit, considering the possibility of generating valuable data automatically. In addition, it would increase independence when making decisions or solving problems without having to resort to an expert in business management.

The ADAGIO project was born in this context. It is a R&D project that combines Big Data and machine learning (ML) strategies for the treatment of geolocated data extracted from heterogeneous data sources. It enables the aggregation, consolidation, and normalization of data from

different semantic fields obtained from the sources mentioned before. Its purpose is to allow reconciled information to be consulted using specific variables, thus facilitating the generation of knowledge.

The application of classification and recommendation systems in this project is of great interest for the interrelation and periodic consolidation of the data process so that the system develops capabilities for transformation, interrelation, and integration of data through supervised learning. In addition, these systems provide a great value for the management of queries, to enhance the performance of queries by users in a language as natural and high level as possible. The fact that the user obtains good results during the searches in the ADAGIO platform is one of the main objectives of the project. In order to improve the user's experience, suggestions are proposed during the phase of filling the search parameters. For this phase, the collaboration of the system users will also be required, evaluating the results of the searches according to their quality and precision.

This study has been performed to facilitate researchers and practitioners the task of choosing the most appropriate system, technology, or algorithm to include in the ADAGIO project for satisfying their requirements. In this sense, this paper presents a systematic mapping study (SMS) that analyzes the current state of the art of the recommendation and classification systems and how they work together. Then, from the point of view of the software development life cycle, this review also shows that the work being done in the ML (for classification and recommendation) research and industrial environment is far from earlier stages such as business requirements and analysis. This makes it very difficult to find efficient and effective solutions that support real business needs from an early stage. Then, this paper suggests the development of new ML research lines to facilitate its application in the different domains.

This paper is organized as follows. Section 2 describes the closest related work to our proposal; Section 3 details the selected method to carry out the SMS; Sections 4 to 8 illustrate the execution of the different phases of the SMS; and finally, Section 9 summarizes the conclusions obtained from the study and presents a set of future work.

2. Related Work

Recommendation and classification systems are acquiring much interest within the scientific community. In this section, the closest related works to the research proposed in this article are presented.

Jaysri et al. [1] presented a complete review of the recommendation systems, focusing on the collaborative filtering. It shows different algorithms based on this filtering for both the user profile and the product characteristics. In addition, it demonstrates several classification methods that may be part of the input for recommendation systems. Ekstrand et al. [2] presented a general overview and focused on the field of recommendation systems. Their purpose was to learn more about the current development of

recommendation methods, specially systems making use of the collaborative filtering.

Obtaining a research perspective on how to make decisions when choosing algorithms to propose recommendations can be found in the paper presented by Gunawardana and Shani [3]. It criticizes the use of online methods, which can offer measures to choose recommendation algorithms, and determines as a crucial element the use of offline tools to obtain these measures. In addition, it discards the use of traditional metrics to make the algorithm choice and reviews the proper elaboration of experiments to carry it out. To do this, the authors perform an analysis of important tasks of the recommendation systems and classify a set of appropriate and well-known assessment measures for each task.

Poussevin et al. [4] exposed the challenge of considering the preferences of users when recommending. The authors analyzed a combination of recommendation systems and classifiers that highlight words that indicate a gap between users' expectations and their actual experience. They conclude that traditional recommendation systems analyze the past classifications; that is, they consider the users' preferences history, while the recommendation systems that analyze the opinion classifications consider the existing evaluations at that moment.

Within the scope of ML, there has been an increase in the interest of the research community, being the subject of many papers. Some of the proposals use lexical classifiers to detect possible feelings using content-based recommendations [5]. Other authors have focused on more traditional branches of ML, using well-known and proven statistical methods such as logistic regression, the Pearson correlation coefficient, or the application of the naive Bayes theorem based on probability, among others [6]. The authors of this paper focused on making extensions of these methods to solve problems inherent in recommendation systems such as cold start or scalability. The cold start [7] is a typical problem since the beginning of the recommendation systems because when a system does not have enough data, precision cannot be assured when recommending. This is a problem that gets worse at the beginning of the implementation of a system when data are not available. Scalability becomes a quite difficult task due to the increase of information in recent years and the amount of data that systems must manage. Recommendation systems, both product and user-based, affect performance and accuracy when these amounts of data are very large. The work presented by Ghazanfar and Prügel-Bennett [8] has been also focused on this problem, generally for the user-based recommendation, which is the most used.

Alternative interesting related work focused in the use of ML is the survey in sentiment classification presented by Hailong et al. [9]. In this work, the authors also provide a comparative study of the techniques found, concluding that supervised ML present a higher accuracy, while lexicon-based methods are likewise competitive because they require less effort and they are not sensitive to the quantity and quality of the training dataset. The survey presented by Mu [10] delivers a review of deep learning-based recommender systems. The authors conclude this work summarizing a set

of future research lines such as cross domain, scalability, explainability, or deep composite model-based recommender systems, among others.

The paper presented by Portugal et al. [11] presents a systematic review of the use of ML in recommender systems. The authors analyzed 121 primary studies classified in different categories: content-based and neighbor-based of content-based filtering, neighborhood-based and model-based of collaborative filtering, and hybrid filtering. This work helps developers to recognize the algorithms, their types, and trends in the use of specific algorithms. It also offers current-type evaluation metrics and categorizes the algorithms based on these metrics. Ouhbi et al. [12] proposed a deep learning-based recommender system to overcome some limitations of existing approaches. In the related work section of this paper, the authors describe a small state of the art of deep learning-based recommender systems, detailing the method, approach, metric, dataset, advantages, and disadvantages of seven proposals.

Zhang et al. [13] delivered a wide review of deep learning-based recommender systems, proposing a classification and highlighting a group of the most influential. The authors debate the pros and cons of using deep learning techniques for recommendation tasks. Additionally, some of the most pressing open problems and promising future extensions are detailed.

In summary, the literature review presented different topics, which may come close to the objective pursued. But there are several differences between these papers and the one presented in this work: (i) the review process: unlike the rest of the papers, this research presents a systematic and rigorous process, ensuring the quality of the results obtained; (ii) the context of application: usually reviews are carried out on the scientific literature; in this case, this research also presents a review on the industrial scope, analyzing the main existing solutions to the problem; and (iii) the scope of application: in this systematic review, the state of the art of the classification and recommendation systems is presented working together, something that in the related works already mentioned is not carried out or it is done independently for classification or recommendation.

3. Methodology

A systematic literature review is an effective way of knowing the state of the art of a subject. This procedure ensures a certain level of quality of information and has the support of the research community. The monitoring of a systematic and guided process guarantees reliable and interesting results and facilitates the work of gathering information.

The review presented in this paper is placed within the context of the recommendation and classification systems from two perspectives: scientific and industrial.

When carrying out a systematic literature review (SLR), the main methodology to be considered is the one presented by Kitchenham and Charters [14]. This is one of the most widely accepted methods in the area of software engineering. It offers a way of performing a SLR consisting in three phases: planning and conducting the review and reporting of

results. However, instead of performing a deep review of the papers comparing them, which is the main goal of a SLR, this study seeks to provide an overview of an interesting topic and to identify the number and type of published-related researches, as well as the related results available. Therefore, the best methodology to be applied is the systematic mapping study (SMS) presented by Petersen et al. [15], a type of the systematic review but with a broader objective. This method will allow identifying the subjects that lack empirical evidence and which are necessary to carry out more empirical studies. SMSs show many similarities with respect to the SLRs. As possible to see in activity diagram of Figure 1, this method establishes a set of five steps, where each of them produces an output. These steps are as follows:

- (i) *Definition of the Research Questions*. Formulation of the research questions (RQs) that will guide the work.
- (ii) *Conduct Search*. The search is normally executed in different digital libraries and based on some keywords extracted from the RQs.
- (iii) *Screening of Papers*. Applying the inclusion and exclusion criteria with the aim of selecting the most relevant and close papers to the topic of the research.
- (iv) *Keywording Using Abstract*. Building of the classification scheme, where all the primary papers selected in the previous phase will be categorized.
- (v) *Data Extraction and Mapping Process*. Data extraction and mapping process based on the results obtained in the keywording activity. This activity will let the researchers to classify which is the state of the art of the topic and to identify gaps and possibilities for future research.

4. Definition of Research Questions

A Research Question (RQ) is the fundamental core of a research project, study, or literature review. Therefore, to know and better understand the existing literature related to the recommendation and classifications systems, it is necessary to formulate a set of research questions. These questions will focus the study, will determine the methodology that will be established, and will guide all the stages of this research. In this sense, the RQs that have been proposed for this SMS are as follows:

- (i) *RQ1*. Which recommendation and classification systems have been researched?
- (ii) *RQ2*. Which recommendation and classification systems have been used?
- (iii) *RQ3*. Which is the nature of the systems found?
- (iv) *RQ4*. Which are the objectives pursued in the proposals found?

5. Conduct Search

Before performing the search in the different digital libraries, it is necessary to complete two operations: define the digital

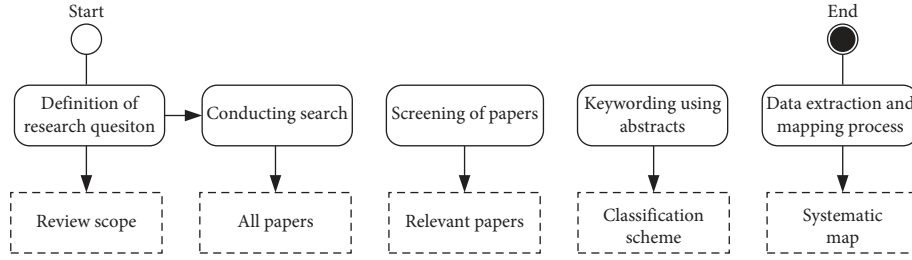


FIGURE 1: SMS workflow.

libraries where the searches will be executed and establish the keywords that will compose the search strings. Selected digital libraries to carry out the search have been the following: SCOPUS, IEEE Xplore, ACM, and ScienceDirect. In addition, for the industrial scope, the search engines that have been selected are Google, Yahoo, and Bing.

To specify the search, keywords were defined, and it is a fundamental part when creating the queries for each digital library. These keywords were obtained after carrying out an analysis of the field of study to which this research applies, recommendation and classification systems. Table 1 shows the complete set of keywords used, and equation (1) shows the formula applied to these keywords to create the final queries.

Boolean expression of keywords is as follows:

$$E_1 = (V_{i=1}^3 A_i) \wedge (V_{j=1}^4 B_j) \wedge (V_{k=1}^4 C_k). \quad (1)$$

Once all the keywords were defined, the queries were constructed. These queries were different for each digital library, and they had different boundary characteristics, depending on the possibilities of the digital library. Digital libraries have certain limitations when conducting searches. For example, some of them do not allow the use of complete search strings; in others, it is necessary to complement these strings with simple textual searches. For this reason, there is the need to create individual queries for each library and, subsequently, to treat the search results to obtain the same results that could have been obtained using the originally proposed query. Table 2 shows a set of examples for each of digital library.

The search was executed on the title, abstract, and keywords of the papers, except in those digital libraries that did not allow it. In such cases, the search was performed on the complete text. Search strings, metadata of found elements (title, author, and year of publication), and summaries of the documents were stored for each search source. Once the first search was executed, it obtained an initial set of 1,195 potential primary studies.

6. Screening of Papers

There are different metrics to define the quality criteria that make a paper relevant. In this work, in addition to those related to the structure of the papers, the quality assurance criteria defined by those scientific papers found that were classified in the following accepted indexes:

- (i) “Journal Citation Report (JCR)” [16] part of the company Thomson Scientific

TABLE 1: Keyword definition.

A	B	C
A1. Machine learning	B1. Recommender system	C1. Classifier
A2. Deep learning	B3. Recommended	C2. Classification
A3. Neuronal networks	B4. Content-based filtering	C3. Classified
	B5. Collaborative filtering	C4. Classification system

TABLE 2: Example of queries.

Digital library	Query
Digital library	TITLE-ABS-KEY (“machine learning” OR “deep learning” OR “neuronal networks”) AND TITLE-ABS-KEY (“classifier” OR “classification” OR “classified” OR “classification system”) AND TITLE-ABS-KEY (“recommended” OR “recommender system” OR “collaborative filtering” OR “content-based filtering”)
Scopus	(“Machine learning” OR “deep learning” OR “neuronal networks”) AND (“classifier” OR “classification” OR “classified” OR “classification system”) AND (“recommended” OR “recommender system” OR “collaborative filtering” OR “content-based filtering”)
IEEE Xplore	acmdlTitle: (“machine learning” “deep learning” “neuronal networks”) OR recordAbstract: (“machine learning” “deep learning” “neuronal networks”) AND (acmdlTitle: (“classifier” “classification” “classified” “classification system”) OR recordAbstract: (“classifier” “classification” “classified” “classification system”)) AND (acmdlTitle: (“recommended” “recommender system” “collaborative filtering” “content-based filtering”) OR recordAbstract: (“recommended” “recommender system” “collaborative filtering” “content-based filtering”))
ACM	(“Machine learning” OR “deep learning” OR “neuronal networks”) AND (“classifier” OR “classification” OR “classified” OR “classification system”) AND (“recommended” OR “recommender system” OR “collaborative filtering” OR “content-based filtering”)
Science direct	(“Machine learning” OR “deep learning” OR “neuronal networks”) AND (“classifier” OR “classification” OR “classified” OR “classification system”) AND (“recommended” OR “recommender system” OR “collaborative filtering” OR “content-based filtering”)

- (ii) The Australian classification created by the “Computing Research and Education Association of Australasia (CORE)” [17]

- (iii) The ranking of relevant congresses for the Scientific Information Society of Spain (SCIE) [18], advising the use of the ranking developed by the Italian associations GII and GRIN [19]

In addition, the following inclusion and exclusion criteria were defined for including or being not a publication into the selected primary studies:

- (i) *C1, Criterion 1.* The classification of the publication in question must be “Computer Science”
- (ii) *C2, Criterion 2.* Written in English
- (iii) *C3, Criterion 3.* The research must be related to the classification and recommendation of data using machine learning systems
- (iv) *C4, Criterion 4.* Searches cannot be repeated. Multiple appearances must be eliminated
- (v) *C5, Criterion 5.* As mentioned above, papers must be classified into the JCR or SCIE rankings
- (vi) *C6, Criterion 6.* The reading of the abstract must fit with the dealt topic

Finally, some recommendations from experts in the subject dealt with in this SMS have also been considered. If these studies were not found after the execution of the different searches, they were included in the final selection of primary studies.

Once defined the quality and inclusion and exclusion criteria, the screening of the papers was performed. According to the C1 of inclusion/exclusion of papers which scope is related to “Computer science,” a total of 923 results were obtained, having discarded 272 papers that did not meet this criterion. C2 was applied to the 923 papers obtained from C1 resulting on 909 papers. To the results obtained from C2, C3 criterion was applied leaving a total of 432 results. Once C4 was applied, a total of 96 papers were removed remaining 336. A total of 259 papers was the result of applying C5. The last filter, C6, was applied resulting on 99 papers considering that 160 of the removed ones did not fit the topic of this research. Finally, repeated papers were removed. This process ended up removing duplicated entries between the different digital libraries.

The result of applying all the quality and inclusion and exclusion criteria was a total of 80 primary studies which will be categorized into the classification schema. The number of papers found corresponds (roughly) to 6% of the results found in the first search. Table 3 shows the primary studies selected.

Figure 2 shows the list of keywords discovered in the different primary studies. In this figure, the keywords are classified based on the total number of matches found between all these primary studies.

Figure 3 depicts the complete process of selecting primary studies. It shows the search procedure for each digital library and the results after the application of each quality and inclusion and exclusion criteria.

By the same token, the process carried out previously was executed for the industrial scope for detecting and

TABLE 3: Selected primary studies.

Title	Reference
Building accurate and practical recommender system algorithms using machine learning classifier and collaborative filtering	[20]
DGA botnet detection using collaborative filtering and density-based clustering	[21]
A multistage collaborative filtering method for fall detection	[22]
Analysis and performance of collaborative filtering and classification algorithms	[1]
Extracting a vocabulary of surprise by collaborative filtering mixture and analysis of feelings	[4]
Content based filtering in online social network using inference algorithm	[23]
Building switching hybrid recommender system using machine learning classifiers and collaborative filtering	[8]
Imputation-boosted collaborative filtering using machine learning classifiers	[24]
CRISP—an interruption management algorithm based on collaborative filtering	[25]
A credit scoring model based on collaborative filtering	[26]
Collaborative filtering recommender systems	[2]
An improved switching hybrid recommender system using naive Bayes classifier and collaborative filtering	[6]
Tweet modeling with LSTM recurrent neural networks for hashtag recommendation	[27]
A two-stage cross-domain recommendation for cold start problem in cyber-physical systems	[28]
ELM based imputation-boosted proactive recommender systems	[29]
Twitter-user recommender system using tweets: a content-based approach	[30]
A personalized time-bound activity recommendation system	[31]
Automated content based short text classification for filtering undesired posts on Facebook	[32]
Shilling attack detection in collaborative recommender systems using a meta learning strategy	[33]
Building a distributed generic recommender using scalable data mining library	[34]
Context-aware movie recommendation based on signal processing and machine learning	[35]
Recommender systems using linear classifiers	[36]
A survey of accuracy evaluation metrics of recommendation tasks	[3]
Incorporating user control into recommender systems based on naive Bayesian classification	[37]
Classification features for attack detection in collaborative recommender systems	[38]
Automatic tag recommendation algorithms for social recommender systems	[39]
Optimizing similar item recommendations in a semi-structured marketplace to maximize conversion	[40]
Capturing knowledge of user preferences: ontologies in recommender systems	[41]
Emotion-based music recommendation using supervised learning	[42]
AWESOME—a data warehouse-based system for adaptive website recommendations	[43]

TABLE 3: Continued.

Title	Reference
Lexical and syntactic features selection for an adaptive reading recommendation system based on text complexity	[5]
A smart-device news recommendation technology based on the user click behavior	[44]
Recommendation as link prediction in bipartite graphs: A graph kernel-based machine learning approach	[45]
A novel approach towards context based recommendations using support vector machine methodology	[46]
A smartphone-based activity-aware system for music streaming recommendation	[47]
An app usage recommender system: improving prediction accuracy for both warm and cold start users	[48]
Proposing design recommendations for an intelligent recommender system logging stress	[49]
A recommender system based on implicit feedback for selective dissemination of eBooks	[50]
A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases	[51]
An approach to content based recommender systems using decision list based classification with k-DNF rule set	[52]
Probabilistic approach for QoS-aware recommender system for trustworthy web service selection	[53]
Approach to cold-start problem in recommender systems in the context of web-based education	[54]
Context and intention-awareness in POIs recommender systems	[55]
A collaborative filtering-based re-ranking strategy for search in digital libraries	[56]
Learning users' interests by quality classification in market-based recommender systems	[57]
Mobile content recommendation system for re-visiting user using content-based filtering and client-side user profile	[58]
A hybrid collaborative filtering algorithm based on KNN and gradient boosting	[59]
A scalable collaborative filtering algorithm based on localized preference	[60]
Recommended or not recommended? Review classification through opinion extraction	[61]
Meta-feature based data mining service selection and recommendation using machine learning models	[62]
Personalized channel recommendation deep learning from a switch sequence	[63]
Affective labeling in a content-based recommender system for images	[64]
A novel approach towards context sensitive recommendations based on machine learning methodology	[65]
A distance-based approach for action recommendation	[66]
Ranking and classifying attractiveness of photos in folksonomies	[67]
Consequences of variability in classifier performance estimates	[68]

TABLE 3: Continued.

Title	Reference
Machine learning and lexicon based methods for sentiment classification: a survey	[9]
Machine learning algorithm selection for forecasting behavior of global institutional investors	[69]
Towards rapid interactive machine learning: evaluating tradeoffs of classification without representation	[70]
Towards a method for automatically evolving Bayesian network classifiers	[71]
A machine learning based trust evaluation framework for online social networks	[72]
Automated problem identification: regression vs. classification via evolutionary deep networks	[73]
Empirical evaluation of ranking prediction methods for gene expression data classification	[74]
Inferring contextual preferences using deep auto-encoding	[75]
Automatic recognition of text difficulty from consumers health information	[76]
A hybrid approach for automatic model recommendation	[77]
Learning instance greedily cloning naive Bayes for ranking	[78]
Pairwise-ranking based collaborative recurrent neural networks for clinical event prediction	[79]
Accurate multi-criteria decision making methodology for recommending machine learning algorithm	[80]
A general extensible learning approach for multi-disease recommendations in a telehealth environment	[81]
An efficient recommendation generation using relevant jaccard similarity	[82]
An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction	[83]
Automatic classification of high resolution land cover using a new data weighting procedure: the combination of k -means clustering algorithm and central tendency measures (KMC-CTM)	[84]
Building a hospital referral expert system with a prediction and optimization-based decision support system algorithm	[85]
Classification techniques on computerized systems to predict and/or to detect apnea: a systematic review	[86]
Identification of category associations using a multilabel classifier	[87]
Making use of associative classifiers in order to alleviate typical drawbacks in recommender systems	[88]
S3Mining: a model-driven engineering approach for supporting novice data miners in selecting suitable classifiers	[89]
The use of machine learning algorithms in recommender systems: a systematic review	[11]

selecting the primary technologies or tools that companies offer. The search engines returned multiple results (Table 4), with a total of 21 proposals remaining were potential candidates.

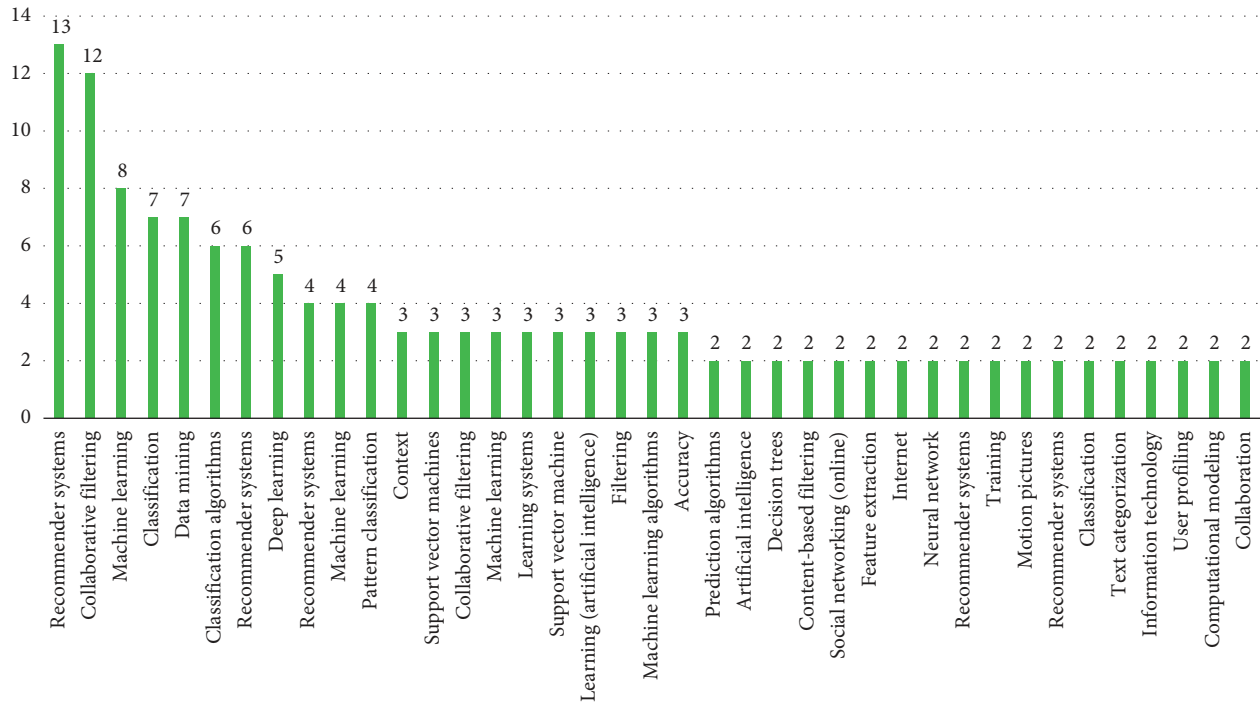


FIGURE 2: Keyword mapping.

7. Keywording using Abstracts

To create the classification scheme for categorizing the selected primary studies, an attempt was made to answer each of the research questions formulated in the planning phase and, in addition, to identify each of them with a set of features.

Moreover, two complete iterations were carried out to classify all the studies and to verify that all the features that had been found included the content of each study. Table 5 shows and describes the classification scheme defined.

Thereupon, process for the definition of the classification scheme is repeated for the industrial area. Through the answer to the research questions and the extraction of the technologies' features, a classification scheme was defined (Table 6).

8. Data Extraction and Mapping Process

8.1. *Scientific Report.* This section describes the most important aspects obtained from the information collected. To achieve this purpose, each of the research questions will be answered and validated, showing the data obtained for each of them. It is important to note that some of the features may appear in several studies; therefore, the totals may not always correspond to 100%.

- (i) Research Question RQ1 finds the methods, techniques, and/or tools that have been investigated for the classification and recommendation systems. Figure 4 shows that the predominant type of studies is methods, which represent 35.00% of the total of the studies, followed by the complete system studies, with a 23.75%. The rest of studies correspond to

algorithms with 20.00%, analysis with a presence of 18.75%, and finally, frameworks with a 6.25% of the total primary studies. From a software development life-cycle perspective (and avoiding methodological discussions), requirements and analysis phases differ from the design phase because it is an earlier stage and closer to the business (or the application model) and is completely technology independent. Then, the found works are contextualized in the technological design phase. No contextualized work was found in early stages (business requirements or analysis).

- (ii) Research Question RQ2 seeks to know the validation of the studies found, which may be practical or theoretical, identifying if they are within the scientific or industrial scope. The results obtained (Figure 5) show that all the primary studies were academic focused. Most of them were validated by some way (97.50%), while 10.00% were not validated.

It is important to note that three different groups have been distinguished within the validation category. The experimentation subgroup includes all those studies whose proposal was tested and validated by experimentation with synthetic and real data sources. This group contains most of the results found that were validated, 72.50% of the total. Another important category is the one that validates the proposals by a case study, which represent 13.75%. Only the 5.00% of the primary studies were carried out through surveys, and just one primary study was focused on the industrial context, representing the 1.25% of the total.

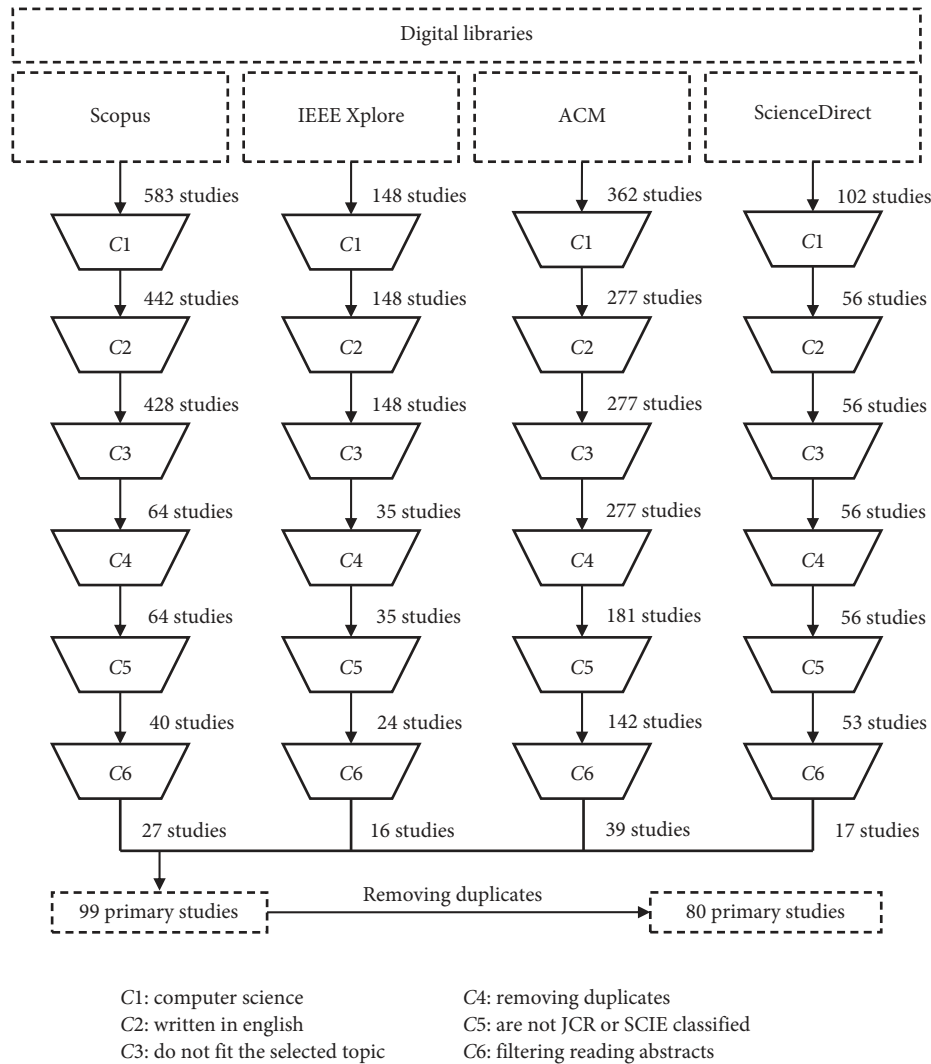


FIGURE 3: Search flow diagram: scientific field.

(iii) Research Question 3 aims to identify the nature of the methods, techniques, and/or tools about the classification and recommendation systems found in the literature. Figure 6 groups two main categories that contain the whole set of features of the primary studies found: recommendation and classification. Within the recommendation group, content-based and collaborative filtering proposals are very balanced, representing the 36.25% and 38.75%, respectively. Hybrid systems are the worst classified with 17.50% of the papers.

Furthermore, the classification group is described, where both supervised and unsupervised learning features are presented. Two features stand out for their use: naive Bayes to classify according to probabilities with a 28.75% and support vectors, representing the 20.00% of total. Target based and Random Forest are the less used, with a presence of just 1 primary study.

(iv) Research Question RQ4 indicates which are the main points of interest of the research and which

areas have been less investigated. This interest is classified into four categories: novelty, analysis, research, and improvement (Figure 7).

The novelty contains those primary studies whose goal is to present something that lacked in the literature, and this category represents 22.50%, with 18 primary studies. Analysis category contains those results that are comparison or study of different existing techniques, and it represents the 7.50% of total. The improvement category represents that 30.00% of the results whose main objective is to improve an existing approach. Finally, the largest category is the research one, where a search on existing or new approaches in the literature is dealt with. It represents the 36.25% of total with 29 primary studies.

At last, it is interesting to analyze other results that are not related to the research questions but with the objective of this document. These results can help to know the evolution of the research of the classification and recommendation systems.

TABLE 4: Selected primary technologies/tools.

Technology	Reference
Scikit-learn	[90]
Surprise	[91]
LightFM	[92]
Rexy	[93]
PredictionIO	[94]
HapiGER	[95]
LensKit	[96]
SuggestGrid	[97]
SLI Systems Recommender	[98]
AmazonWebService Machine Learning	[99]
Azure ML Studio	[100]
Yusp	[101]
IBM Watson	[102]
Recombee	[103]
Mr. DLib	[104]
Caret	[105]
Shiny	[106]
RandomForest	[107]
KlaR	[108]
CORElearn	[109]
RecommenderLab	[110]

TABLE 5: Scientific classification scheme.

Research Question	Feature	Description
RQ1	Algorithm	This feature defines if the primary study proposes an algorithm or series of algorithms
	System	This feature defines if the primary study is a software system based on different components
	Framework	This feature defines whether the primary study is based on a framework
	Method	This feature defines if the primary study is a set of procedures to obtain a result
	Analysis	This feature defines if the primary study is a theoretical study based on surveys or systematic reviews among others
RQ2	Validated	This feature defines whether the primary study has been validated with experiments, use cases, or surveys
	Not validated	This feature defines if the primary study has not been validated with experiments, use cases, or surveys
	Academic	This feature defines if the primary study has been validated with some academic case study
	Industrial	This feature defines if the primary study has been validated with some case study in the industry
	Experiment	This feature defines if the primary study has been validated with the elaboration of different experiments
	Use case	This feature defines whether the primary study has been validated with the study of a use case
	Survey	This feature defines if the primary study has been validated with the elaboration of some type of survey

TABLE 5: Continued.

Research Question	Feature	Description
RQ3	Content based	This feature defines whether the solution proposed by the primary study is based or composed of a recommendation system with a content-based filter
	Collaborative	This feature defines whether the solution proposed by the primary study is based or composed of a recommendation system with a collaborative filter
	Hybrid	This feature defines whether the solution proposed by the primary study is based or composed of a collaborative and content-based filter conjunction
	Graph kernel	This feature defines whether the primary study is based or composed of a graphic classifier
	Naive Bayes	This feature defines whether the primary study is based or composed of a naive probabilistic classifier naive Bayes
	Logistic regression	This feature defines whether the primary study is based or composed of a classifier by logistic regression
	Decision tree	This feature defines whether the primary study is based or composed of a classifier by decision trees
	Lexical	This feature defines whether the primary study is based or composed of a classifier based on textual features
	Based on rules	This feature defines if the primary study is based or composed of a rule-based classifier
	Neural networks	This feature defines if the primary study is based or composed of a classifier based on neural networks
	Clustering	This feature defines if the primary study is based or composed of a non-supervised cluster classifier
	Boosting	This feature defines if the primary study is based or composed of an ensemble classifier with a boosting scheme
	Linear algorithm	This feature defines if the primary study is based on or composed of a classifier based on a linear algorithm
	Based on attributes	This feature defines whether the primary study is based or composed of a classifier based on attributes
	Multiclass	This feature defines if the primary study is based or composed of a multiclass classifier
	Warehouse	This feature defines if the primary study is composed of a classifier based on data warehouse
	SVM vectors	This feature defines if the primary study is composed of a classifier that makes use of support vector machines
	Neighbor method	This feature defines if the primary study is composed of a classifier based on the neighbor method
	Opinion-based	This feature defines if the primary study is composed of an opinion-based classifier
	Target-based	This feature defines if the primary study is composed of a target-based classifier
Random forest	This feature defines if the primary study is composed of a random forest classifier	
RQ4	Novelty	This feature defines if the primary study is a new proposal that does not exist in the literature
	Analysis	This feature defines if the primary study is an analysis of several existing proposals in the literature
	Research	This feature defines if the primary study is an investigation of existing or new proposals
	Improvement	This feature defines if the primary study is an improvement of an existing proposal in the literature

TABLE 6: Industrial classification scheme.

Research Question	Feature	Description
RQ1	Tool	This feature defines if the technology found is a tool for supporting
	Library	This feature defines if the technology found is a library of methods or framework
	System	This feature defines if the technology found is a complete system
	Platform	This feature defines if the technology found is a platform
	API	This feature defines if the technology found is an API that offers its functionalities
RQ2	Free	This feature defines whether the technology found is free software
	Commercial	This feature defines if the technology found is proprietary software
RQ3	Python	This feature defines whether the technology found is based on python
	Apache Spark	This feature defines whether the technology found is based on Apache Spark
	Node	This feature defines whether the technology found is based on node
	Java	This feature defines whether the technology found is based on java
	Ruby	This feature defines whether the technology found is based on ruby
	Unknown	This feature defines if the technology found does not allow knowing in what language it is based
RQ4	Recommendation	This feature defines if the technology found is aimed at the recommendation
	Classification	This feature defines if the technology found is aimed at the classification

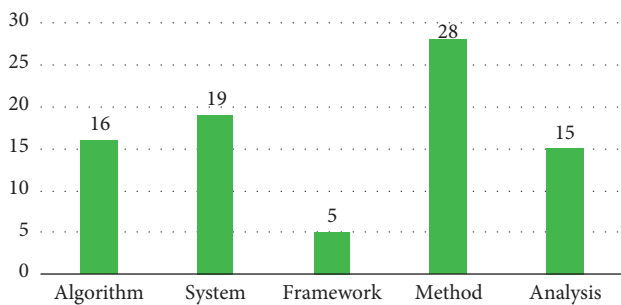


FIGURE 4: Scientific field: Research Question 1.

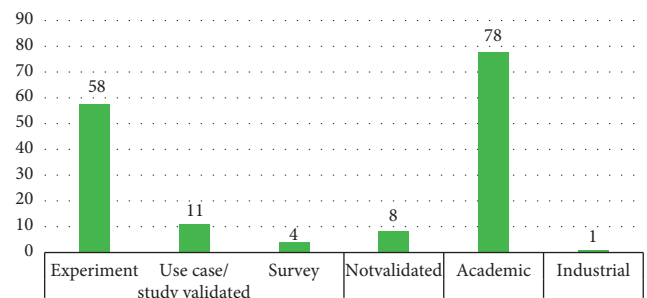


FIGURE 5: Scientific field: Research Question 2.

- (i) Figure 8 shows the trend of publication in topics related to the classification and recommendation systems. The chart shows that the trend increases in recent years, so it can be deduced that it is a subject of high interest to the scientific community. It is important to note that, at the beginning of 2019, there are already more than half of the papers selected for the previous year.
- (ii) Figure 9 presents the number of papers obtained for each of the digital libraries and the relationship with those finally selected for further study. In light green, the initial results are shown, highlighting ACM with 27 papers shown, followed by SCOPUS and IEEE

Xplore with 23 and 14, respectively. ScienceDirect returned only 4 results. Dark green shows the finally selected studies of each digital library.

8.2. *Industrial Report.* After the description of the results obtained from the scientific report, this section presents the report of the data bring about conducting the study of the industrial scope.

- (i) Research Question RQ1 finds the products that have been developed for the classification and recommendation systems. Figure 10 shows that the most frequent results have been complete systems and

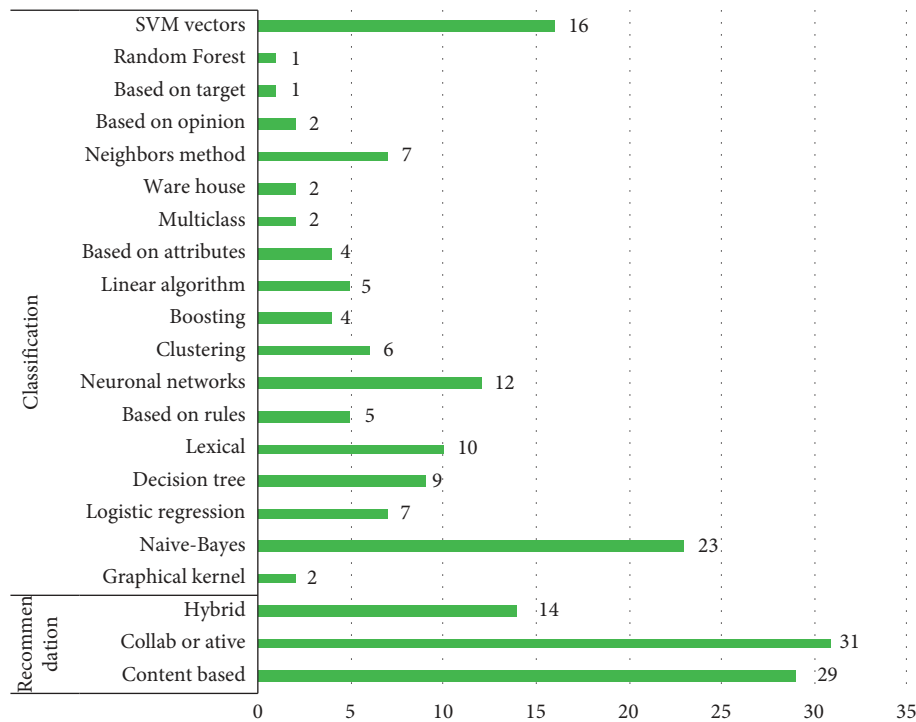


FIGURE 6: Scientific field: Research Question 3.

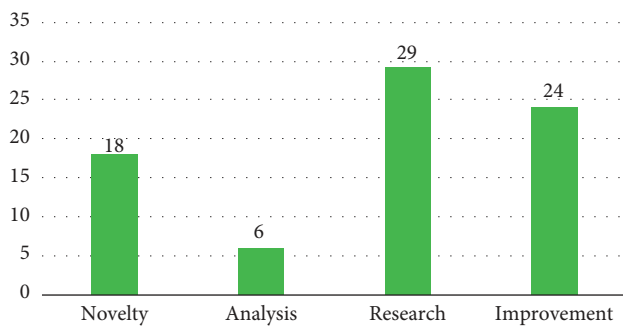


FIGURE 7: Scientific field: Research Question 4.

libraries or frameworks, with 5 and 4 proposals, respectively. The next two features are the APIs and tools, representing 3 and 4 proposals, respectively. In the last place, it located the platform feature, with just one proposal found.

The sum of the complete systems and the libraries represent 47.62% of the total of the proposals. The set of technologies that represent the APIs is 14.29%, the tools 9.52%, and finally, the platform is 4.76% of the total. From a software development life-cycle perspective (and avoiding methodological discussion), requirements and analysis phases differ from the design phase because it is an earlier stages and closer to the business (or application model) and is completely technology independent. Then, the found works are contextualized in the technological design phase. No contextualized work was found in early stages (business requirements or analysis).

Research Question RQ2 aims to determine if the products obtained in this scope are free or proprietary software. This classification has great interest to know those that can suppose an extra cost for the execution of the project.

According to the taxonomy defined, Figure 11 shows that results are balanced to the open side; commercial software, with 8 proposals, represent 38.10% of the total, and the set of free software technologies is composed of 12 results, 57.14% of the total.

- (ii) Research Question RQ3 seeks to identify the nature of the products found. According to the taxonomy carried out after the extraction of features, results obtained are shown in Figure 12. It has been found that there is a group that gathers most of the technologies. This group corresponds to *Python*, with 7 results, representing 33.33% of the total. The next group with the highest results is *R*, with 28.57% after returning 6 results. After that, *Java* is placed, representing the 19.05% of total. Next, *Apache Spark* technology is classified with 3 proposals obtained, 14.29% of the total. Finally, there are two technologies with a single appearance, and they are *Node* and *Ruby*, with 9.52% of the total proposals found. Within this research question, it is highlighted that a large amount of proprietary software did not allow to know what technology they are based on so they were included in the category of others. This category turned out to be 14.29% of the results, with 3 proposals.

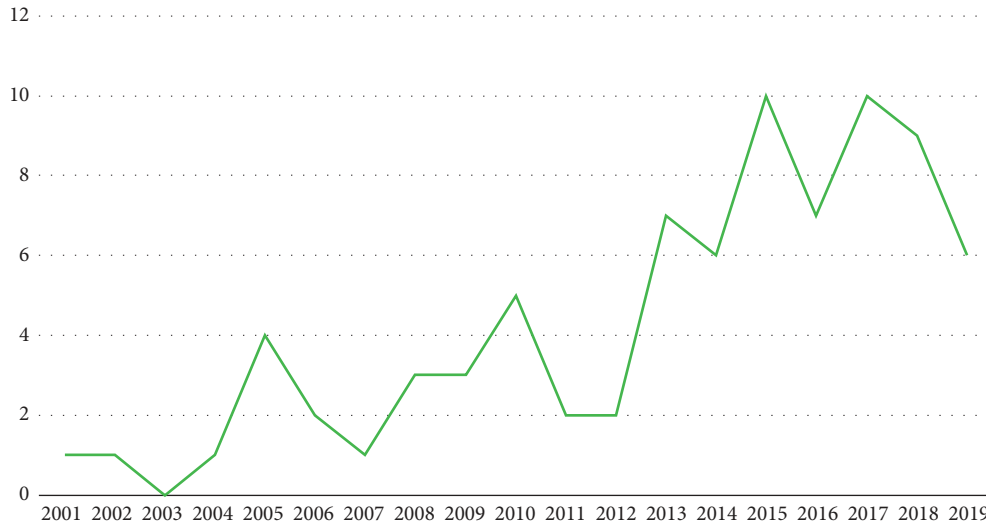


FIGURE 8: Results per year.

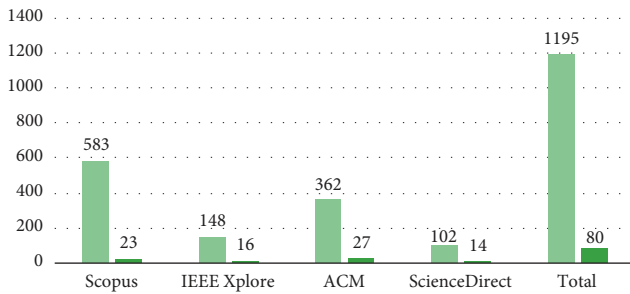


FIGURE 9: Selected by database.

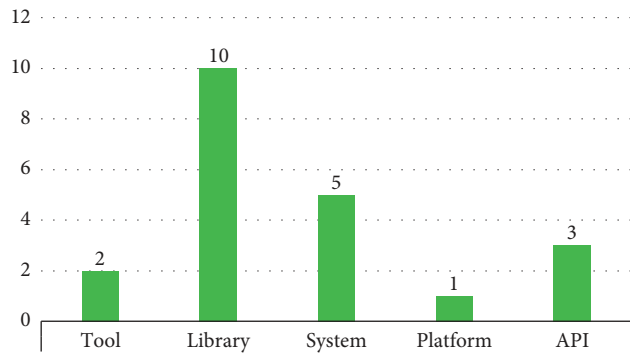


FIGURE 10: Industrial scope: Research Question 1.

(iii) Research Question RQ4 locates the main objective of the technology. In this case, two different groups have been established: classification and recommendation systems (Figure 13). In the case of the technologies that offer a classification system, a total of 10 proposals was obtained, representing 47.62% of the technologies implemented. In the case of recommendation systems, 76.19% of the technologies offered a solution to this problem; that is, 16 of the proposals were found. Finally, it is important to note that the 28.57% (6 proposals) of the total use both regression and classification.

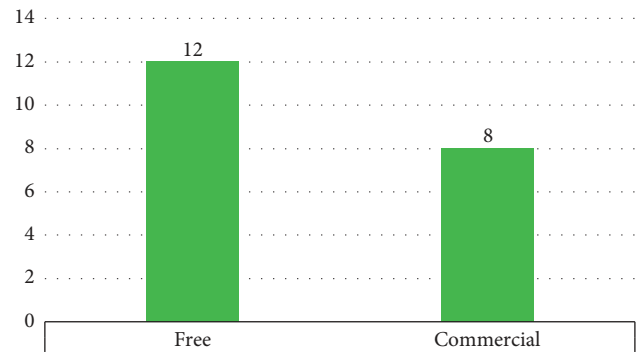


FIGURE 11: Industrial scope: Research Question 2.

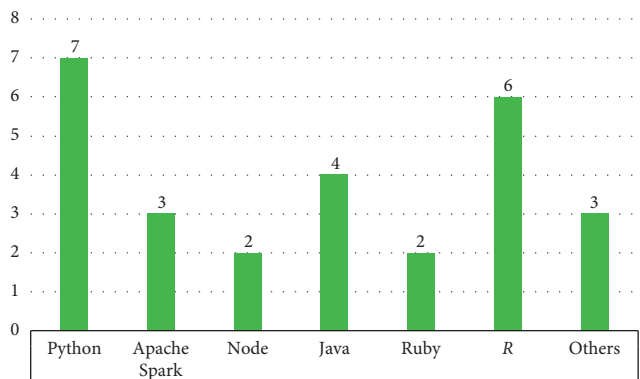


FIGURE 12: Industrial scope: Research Question 3.

9. Conclusion and Future Work

The development of this research has meant an immersion in the depths of the recommendation and classification systems, presenting a SMS which aims to illustrate the state of the art of these systems nowadays. In addition, with the execution of this study, it has been intended to offer help in decision-making about the algorithms to be implemented in the ADAGIO project.



FIGURE 13: Industrial scope: Research Question 4.

Unlike most SMS, that are focused on the scientific literature, and this study has been carried out from two points of view as discussed throughout the paper: the scientific and the industrial scopes.

A total of 80 primary studies obtained from the main digital libraries were analyzed. Within the scientific field, the results showed that the most studied technique in recommendation systems is recommendation with the use of collaborative filters, closely followed by those that use content-based filters. Only 14 used hybrid recommendation systems, whereas 31 used collaborative filtering and 29 used content-based methods. This is an interesting suggestion for researchers starting to use recommender systems, to find which of them are more popular and more used in the scientific environment. As there are more recommender systems than classification models, it seems that recommendation is well known for scientific researchers, and the most used technique is collaborative filtering.

In the case of classification solutions, the most researched alternatives correspond to naive Bayes, SVM vectors, and neuronal networks, representing almost 55% of the techniques used for this purpose. These results are due to the great presence of studies oriented to social networks, which cover a large part of Internet traffic.

It is important to point out that all the studies analyzed in the scientific field were found to be of a theoretical nature; i.e., none of them are within the industrial scope. Although many of the proposals present a validation, few of them use real data sources instead of synthetic ones (artificially generated rather than generated by real-world events) to carry out their experiments. In this sense, a lack of technology transfer of these proposals to real case studies has been detected.

Furthermore, by conducting market research through systematic industrial mapping, it was found that there are many technologies that offer automatic learning solutions, and most of which are complete systems or libraries. However, the nature of most of them could not be known because the proprietary software did not allow it. Another important issue that must be highlighted is that not only the communities of free software developers are interested in this topic but also there are large companies that are working on it for commercial purposes. This clearly shows the underlying economic interest, an indicator that it is a branch of long-distance research.

During the execution of the research on this subject, few studies were discovered that offered improvements to specific problems through the combination of recommendation and classification systems, the main motivation for this work. In the literature analyzed, the most interesting solutions, algorithms, and technologies have been found also to be used independently for classification and regression. This research is not only useful for the researcher trying to use both models at the same time but also for the analysts trying to do just classification or just regression. As future work, a very interesting research line may focus on how to combine these systems to obtain more efficient and effective solutions.

From a software development life-cycle perspective (and avoiding methodological discussions), requirements and analysis phases differ from the design phase because it is an earlier stages and closer to the business (or application model) and is completely technology independent. This SMS shows that the majority of all work carried out in the ML research and industrial field (combining classification and recommendation algorithms) respond to the design and implementation phase but are far from offering solutions in earlier stages such as requirements and analysis. This makes it very difficult to find efficient and effective solutions that support real business needs from an early stage. The present work let justify the opening of new ML research lines to support the information system development since early stages. A hypothetical solution proposal could be to provide business analysts with theoretical frameworks and support tools that facilitate the efficient and effective resolution of problems and that, subsequently, will allow the automation of their design and implementation. Specifically, this solution could consist of the definition of a theoretical framework:

9.1. Foundational Knowledge

- (i) *Archetype Models for the Different Application Domains.* This model is used for the conceptualization, formalization, and categorization of the application domains under study. The objective is to understand which application domains exist and which is the basic information structure that should support the application domain. Through the development of these predefined archetype models, information structures could be offered in a systematic way in order to offer support to the different existing problems.
- (ii) *Classification and Recommendation Template Methods to be Applied to Archetype Models.* This model is used for the conceptualization, formalization, and categorization of ML solutions (combining classification and recommendation algorithms) for all those application domains that have been defined by means of archetype models. The objective is to facilitate the development of a framework that allows the automatic generation of ML solutions and that, in addition, could adjust the classification and the recommendation according to the needs of each application domain.

9.2. Applied Knowledge

- (i) From a strategic point of view, understanding the strategy as a set of ordered stages or phases (phase 1: classification and phase 2: recommendation)

Define ML solution strategies based on the combination of classification algorithms and recommendation. In other words, determine to what extent and in what manner (iterative and iterative-incremental) the classification and recommendation phases should be combined for a more efficient and effective use of these algorithms in problem solving. In addition, the above strategies may depend on the application domain being studied. Determine which strategic configurations are most appropriate for each application domain. The idea is to facilitate decision-making by automating decisions by entering a particular application domain or problem.

- (ii) From a tactical point of view

Determine which machine learning methods, techniques, and tools are the most effective and efficient for the application of the previous strategies, determining the most appropriate for each phase (classification and recommendation) according to the application domain of the object of study.

Finally, we can accomplish that even having executed this rigorous study, there is still a big difficulty in deciding about which algorithm is better than another depending on the context in which it is used. There is no generic classifier or recommender, and several should be implanted depending on the type of data. Currently, it also depends on the desired level of complexity and the cost of misclassification. In conclusion, there is no better model, and everything depends on the characteristics of each problem. In this sense, another possible future work is to characterize these systems, with formal methods (e.g., QuEF [111]), to reduce the cost when making decisions about it.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research has been supported by the Pololas project (TIN2016-76956-C3-2-R) of the Spanish Ministry of Economy and Competitiveness, the ADAGIO (P106-16/E09) project of the Centro para el Desarrollo Tecnológico Industrial (CDTI) of Spain, the Agencia Estatal de Investigación, Spain (Project MTM2017-86875-C3-2-R), and Gobierno de Extremadura, Spain (Project GR18108).

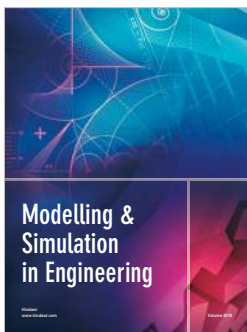
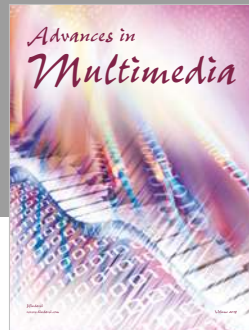
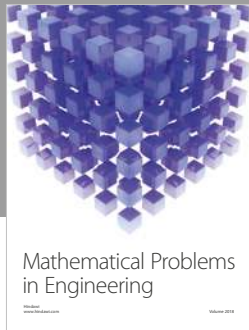
References

- [1] S. Jaysri, J. Priyadharshini, P. Subathra, and Dr. (Col.) P. N. Kumar, "Analysis and performance of collaborative filtering and classification algorithms," *International Journal of Applied Engineering Research*, vol. 10, pp. 24529–24540, 2015.
- [2] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, "Collaborative Filtering Recommender Systems," *Foundations and Trends® in Human—Computer Interaction*, vol. 4, no. 2, pp. 81–173, 2011.
- [3] A. Gunawardana and G. Shani, "A survey of accuracy evaluation metrics of recommendation tasks," *Journal of Machine Learning Research*, pp. 2935–2962, 2009.
- [4] M. Poussevin, V. Guigue, and P. Gallinari, "Extracting a vocabulary of surprise by collaborative filtering mixture and analysis of feelings," in *Proceedings of the CORIA 2015—Conference in Search Informations and Applications—12th French Information Retrieval Conference*, Paris, France, March 2015.
- [5] M. Z. Kurdi, "Lexical and syntactic features selection for an adaptive reading recommendation system based on text complexity," in *Proceedings of the 2017 International Conference on Information System and Data Mining*, pp. 66–69, Charleston, SC, USA, April 2017.
- [6] M. A. Ghazanfar and A. Prügel-Bennett, "An improved switching hybrid recommender system using naive Bayes classifier and collaborative filtering," in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 (IMECS)*, Hong Kong, China, 2010.
- [7] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval—SIGIR '02*, New York, NY, USA, 2002.
- [8] M. Ghazanfar and A. Prügel-Bennett, "Building switching hybrid recommender system using machine learning classifiers and collaborative filtering," *IAENG International Journal of Computer Science*, vol. 37, no. 3, 2010.
- [9] Z. Hailong, G. Wenyan, and J. Bo, "Machine learning and lexicon based methods for sentiment classification: a survey," in *Proceedings of the 11th Web Information System and Application Conference (WISA)*, pp. 262–265, Tianjin, China, September 2014.
- [10] R. Mu, "A survey of recommender systems based on deep learning," *IEEE Access*, vol. 6, pp. 69009–69022, 2018.
- [11] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: a systematic review," *Expert Systems with Applications*, vol. 97, pp. 205–227, 2018.
- [12] B. Ouhbi, B. Frikh, E. Zemmouri, and A. Abbad, "Deep learning based recommender systems," *IEEE International Colloquium on Information Science and Technology (CiSt)*, vol. 2018, pp. 161–166, 2018.
- [13] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: a survey and new perspectives," *ACM Computing Surveys*, vol. 52, no. 1, p. 5, 2019.
- [14] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *Engineering*, vol. 2, p. 1051, 2007.
- [15] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*, vol. 17, p. 10, Bari, Italy, 2008.
- [16] L. Leydesdorff, "Top-down decomposition of the journal citation report of the social science citation index: graph- and factor-analytical approaches," *Scientometrics*, vol. 60, no. 2, pp. 159–180, 2004.

- [17] J. L. C. Izquierdo, V. Cosentino, and J. Cabot, "Analysis of co-authorship graphs of CORE-ranked software conferences," *Scientometrics*, vol. 109, no. 3, pp. 1665–1693, 2016.
- [18] SCIE, "La Sociedad Científica Informática de España," 2017.
- [19] SCIE, "GII-GRIN-SCIE (GGS) Conference Rating," 2019.
- [20] A. Sattar, M. A. Ghazanfar, and M. Iqbal, "Building accurate and practical recommender system algorithms using machine learning classifier and collaborative filtering," *Arabian Journal for Science and Engineering*, vol. 42, no. 8, pp. 3229–3247, 2017.
- [21] T.-D. Nguyen, T.-D. Cao, and L.-G. Nguyen, "DGA botnet detection using collaborative filtering and density-based clustering," in *Proceedings of the Sixth International Symposium on Information and Communication Technology*, pp. 203–209, Hue City, Vietnam, December 2015.
- [22] T. Xie, Y. Chen, L. Hu, C. Gao, C. Hu, and J. Shen, "A multistage collaborative filtering method for fall detection," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, Rio, Brazil, August 2017.
- [23] N. Thilagavathi and R. Taarika, "Content based filtering in online social network using inference algorithm," in *Proceedings of the 2014 International Conference on Circuits, Power and Computing Technologies (ICCPCT)*, Nagercoil, India, March 2014.
- [24] X. Su, T. M. Khoshgoftaar, X. Zhu, and R. Greiner, "Imputation-boosted collaborative filtering using machine learning classifiers," in *Proceedings of the 2008 ACM Symposium on Applied Computing—SAC '08*, Fortaleza, Ceará, Brazil, March 2008.
- [25] T. Shrot, A. Rosenfeld, J. Golbeck, and S. Kraus, "CRISP -an interruption management algorithm based on collaborative filtering," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, Canada, 2014.
- [26] X. Zheng, "A credit scoring model based on collaborative filtering," in *Proceedings of the 9th International Conference on Computational Intelligence and Security*, Emei Mountain, Sichuan, China, December 2013.
- [27] J. Li, H. Xu, X. He, J. Deng, and X. Sun, "Tweet modeling with LSTM recurrent neural networks for hashtag recommendation," in *Proceedings of the International Joint Conference on Neural Networks*, Vancouver, British Columbia, Canada, 2016.
- [28] P. Liu, J. Cao, X. Liang, and W. Li, "A two-stage cross-domain recommendation for cold start problem in cyber-physical systems," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 2015.
- [29] P. Bedi, Richa, S. K. Agarwal, and V. Bhasin, "ELM based imputation-boosted proactive recommender systems," in *Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Jaipur, India, September 2016.
- [30] R. H. Nidhi and B. Annappa, "Twitter-user recommender system using tweets: a content-based approach," in *Proceedings of the ICCIDS 2017 International Conference on Computational Intelligence in Data Science*, pp. 1–6, Chennai, India, June 2017.
- [31] R. Mittal and V. Sinha, "A personalized time-bound activity recommendation system," in *Proceedings of the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, USA, January 2017.
- [32] A. S. Vairagade and R. A. Fadnavis, "Automated content based short text classification for filtering undesired posts on Facebook," in *Proceedings of the IEEE World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCTFTR)*, Coimbatore, India, 2016.
- [33] W. Bhebe and O. P. Kogeda, "Shilling attack detection in collaborative recommender systems using a meta learning strategy," in *Proceedings of the 2015 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, pp. 56–61, IEEE, Windhoek, Namibia, May 2015.
- [34] L. Bhatia and S. S. Prasad, "Building a distributed generic recommender using scalable data mining library," in *Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Communication Technology (CICT)*, Ghaziabad, India, 2015.
- [35] C. Biancalana, F. Gasparetti, A. Micarelli, A. Miola, and G. Sansonetti, "Context-aware movie recommendation based on signal processing and machine learning," in *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, Chicago, IL, USA, 2011.
- [36] T. Zhang and V. S. Iyengar, "Recommender systems using linear classifiers," *Journal of Machine Learning Research*, pp. 313–334, 2002.
- [37] V. Pronk, W. Verhaegh, A. Proidl, and M. Tiemann, "Incorporating user control into recommender systems based on naive Bayesian classification," in *Proceedings of the ACM International Conference on Recommender Systems*, Minneapolis, MN, USA, 2007.
- [38] R. Burke, B. Mobasher, C. Williams, and R. Bhaumik, "Classification features for attack detection in collaborative recommender systems," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '06*, Philadelphia, PA, USA, August 2006.
- [39] Y. Song, L. Zhang, and C. L. Giles, "Automatic tag recommendation algorithms for social recommender systems," *ACM Transactions on the Web*, vol. 5, no. 1, p. 31, 2011.
- [40] Y. M. Brovman, "Optimizing similar item recommendations in a semi-structured marketplace to maximize conversion," in *Proceedings of the 10th ACM Conference on Recommender Systems—RecSys '16*, Boston, MA, USA, September 2016.
- [41] S. E. Middleton, D. C. De Roure, and N. R. Shadbolt, "Capturing knowledge of user preferences," in *Proceedings of the International Conference on Knowledge capture—K-CAP*, Victoria, BC, Canada, 2001.
- [42] P. P. Jean-Jacques, J. Noack, and K. Bodarwé, "Emotion-based music recommendation using supervised learning," in *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, Linz, Austria, December 2015.
- [43] A. Thor and E. Rahm, "AWESOME—A Data Warehouse-Based System for Adaptive Website Recommendations," in *Proceedings of the Thirtieth International Conference on Very Large Data Bases*, vol. 30, pp. 384–395, VLDB Endowment, Toronto, Ontario, Canada, September 2004.
- [44] Y. H. Gu, S. J. Yoo, Z. Piao, J. No, Z. Jiang, and H. Yin, "A smart-device news recommendation technology based on the user click behavior," in *Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory*, pp. 9–16, Jeju Island, Republic of Korea, October 2016.
- [45] X. Li and H. Chen, "Recommendation as link prediction in bipartite graphs: a graph kernel-based machine learning approach," *Decision Support Systems*, vol. 54, no. 2, pp. 880–890, 2013.
- [46] A. A. Kothari and W. D. Patel, "A novel approach towards context based recommendations using support vector

- machine methodology,” *Procedia Computer Science*, vol. 57, pp. 1171–1178, 2015.
- [47] W. P. Lee, C. T. Chen, J. Y. Huang, and J. Y. Liang, “A smartphone-based activity-aware system for music streaming recommendation,” *Knowledge-Based Systems*, vol. 131, pp. 70–82, 2017.
- [48] D. Han, J. Li, W. Li, R. Liu, and H. Chen, *An App Usage Recommender System: Improving Prediction Accuracy for Both Warm and Cold Start Users*, Multimedia Systems, 2019.
- [49] A. Visuri, R. Poguntke, and E. Kuosmanen, *Proposing Design Recommendations for an Intelligent Recommender System Logging Stress*, Association for Computing Machinery, New York, NY, USA, 2018.
- [50] E. R. Núñez-Valdez, D. Quintana, R. G. Crespo, P. Isasi, and E. Herrera-Viedma, “A recommender system based on implicit feedback for selective dissemination of ebooks,” *Information Sciences*, vol. 467, pp. 87–98, 2018.
- [51] S. Narayan and E. Sathiyamoorthy, “A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases,” *Neural Computing and Applications*, vol. 31, no. S1, pp. 93–102, 2019.
- [52] A. Pujahari and V. Padmanabhan, “An approach to content based recommender systems using decision list based classification with k-DNF rule set,” in *Proceedings of the 2014 13th International Conference on Information Technology (ICIT)*, Bhubaneswar, India, December 2014.
- [53] M. Mehdi, N. Bouguila, and J. Bentahar, “Probabilistic approach for QoS-aware recommender system for trustworthy web service selection,” *Applied Intelligence*, vol. 41, no. 2, pp. 503–524, 2014.
- [54] R. A. Gotardo, E. R. Hruschka, S. D. Zorzo, and P. R. M. Cereda, “Approach to cold-start problem in recommender systems in the context of web-based education,” in *Proceedings of the 2013 12th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA, December 2013.
- [55] H. Costa, B. Furtado, D. Pires, L. Macedo, and A. Cardoso, “Context and intention-awareness in POIs recommender systems,” in *Proceedings of the 6th ACM Recommender Systems Conference, 4th Workshop on Context-Aware Recommender Systems (RecSys)*, vol. 12, p. 5, Dubai, UAE, September 2012.
- [56] U. Rohini and V. Ambati, “A collaborative filtering based re-ranking strategy for search in digital libraries,” in *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2005.
- [57] Y. Z. Wei, L. Moreau, and N. R. Jennings, “Learning users’ interests by quality classification in market-based recommender systems,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 12, pp. 1678–1688, 2005.
- [58] W. Paireekreng, “Mobile content recommendation system for re-visiting user using content-based filtering and client-side user profile,” in *Proceedings—International Conference on Machine Learning and Cybernetics*, Lanzhou, China, 2013.
- [59] S. Lu, B. Wang, H. Wang, and Q. Hong, “A hybrid collaborative filtering algorithm based on KNN and gradient boosting,” in *Proceedings of the 13th International Conference on Computer Science and Education (ICCSE)*, Colombo, Sri Lanka, August 2018.
- [60] L. Zhang, B. Xiao, J. Guo, and C. Zhu, “A scalable collaborative filtering algorithm based on localized preference,” in *Proceedings of the 7th International Conference on Machine Learning and Cybernetics (ICMLC)*, Melbourne, Australia, December 2008.
- [61] S. Feng, M. Zhang, Y. Zhang, and Z. Deng, “Recommended or not recommended? Review classification through opinion extraction,” in *Proceedings of the 12th Asia-Pacific Web Conference, Advances in Web Technologies and Applications (APWeb)*, Busan, Korea, April 2010.
- [62] B. Alghofaily and C. Ding, “Meta-feature based data mining service selection and recommendation using machine learning models,” in *Proceedings of the 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, Xi’an, China, October 2018.
- [63] C. Yang, S. Ren, Y. Liu, H. Cao, Q. Yuan, and G. Han, “Personalized channel recommendation deep learning from a switch sequence,” *IEEE Access*, vol. 6, pp. 50824–50838, 2018.
- [64] M. Tkalčič, A. Odić, A. KoTkalšičir, and J. Tasič, “Affective labeling in a content-based recommender system for images,” *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 391–400, 2013.
- [65] A. A. Kothari and W. D. Patel, “A novel approach towards context sensitive recommendations based on machine learning methodology,” in *Proceedings of the 2015 5th International Conference on Communication Systems and Network Technologies (CSNT)*, Gwalior, MP, India, April 2015.
- [66] R. Trepos, A. Salleb, M. O. Cordier, V. Masson, and C. Gascuel, “A distance-based approach for action recommendation,” in *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2005.
- [67] J. S. Pedro and S. Siersdorfer, “Ranking and Classifying Attractiveness of Photos in Folksonomies,” in *Proceedings of the 18th International Conference on World Wide Web*, pp. 771–780, ACM, Madrid, Spain, April 2009.
- [68] T. Raeder, T. R. Hoens, and N. V. Chawla, “Consequences of variability in classifier performance estimates,” in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, Sydney, Australia, 2010.
- [69] J. J. Ahn, S. J. Lee, K. J. Oh, T. Y. Kim, H. Y. Lee, and M. S. Kim, “Machine learning algorithm selection for forecasting behavior of global institutional investors,” in *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences (HICSS)*, Waikoloa, Hawaii, January 2009.
- [70] D. Arendt, E. Saldanha, R. Wesslen, S. Volkova, and W. Dou, “Towards rapid interactive machine learning: evaluating tradeoffs of classification without representation,” in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 591–602, Marina del Ray, CA, USA, March 2019.
- [71] A. G. C. de Sá and G. L. Pappa, “Towards a method for automatically evolving bayesian network classifiers,” in *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, pp. 1505–1512, ACM, Amsterdam, Netherlands, July 2013.
- [72] K. Zhao and L. Pan, “A machine learning based trust evaluation framework for online social networks,” in *Proceedings of the 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, Beijing, China, September 2014.
- [73] E. Dufourq and B. A. Bassett, “Automated problem identification: regression vs. classification via evolutionary deep networks,” in *Proceedings of the South African Institute of Computer Scientists and Information Technologists*, p. 12, Thaba Nchu, South Africa, September 2017.

- [74] B. F. De Souza, A. C. P. L. F. De Carvalho, and C. Soares, "Empirical evaluation of ranking prediction methods for gene expression data classification," in *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2010.
- [75] M. Unger, B. Shapira, L. Rokach, and A. Bar, "Inferring contextual preferences using deep auto-encoding," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, pp. 221–229, ACM, Bratislava, Slovakia, July 2017.
- [76] W. Yunli, "Automatic recognition of text difficulty from consumers health information," in *Proceedings of the IEEE Symposium on Computer-Based Medical Systems*, Salt Lake City, Utah, 2006.
- [77] R. Vainshtein, A. Greenstein-Messica, G. Katz, B. Shapira, and L. Rokach, "A hybrid approach for automatic model recommendation," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 1623–1626, ACM, Turin, Italy, October 2018.
- [78] L. Jiang and H. Zhang, "Learning instance greedily cloning naive Bayes for ranking," in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, p. 8, IEEE, Houston, TX, USA, 2005.
- [79] Z. Qiao, S. Zhao, C. Xiao, X. Li, Y. Qin, and F. Wang, "Pairwise-ranking based collaborative recurrent neural networks for clinical event prediction," in *Proceedings of the IJCAI International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, July 2018.
- [80] R. Ali, S. Lee, and T. C. Chung, "Accurate multi-criteria decision making methodology for recommending machine learning algorithm," *Expert Systems with Applications*, vol. 71, pp. 257–278, 2017.
- [81] R. Lafta, J. Zhang, X. Tao et al., "A general extensible learning approach for multi-disease recommendations in a telehealth environment," *Pattern Recognition Letters*, 2018.
- [82] S. Bag, S. K. Kumar, and M. K. Tiwari, "An efficient recommendation generation using relevant jaccard similarity," *Information Sciences*, vol. 483, pp. 53–64, 2019.
- [83] A. Soudani and W. Barhoumi, "An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction," *Expert Systems with Applications*, vol. 118, pp. 400–410, 2019.
- [84] S. S. Durduran, "Automatic classification of high resolution land cover using a new data weighting procedure: the combination of k -means clustering algorithm and central tendency measures (KMC-CTM)," *Applied Soft Computing*, vol. 35, pp. 136–150, 2015.
- [85] C. L. Chi, W. N. Street, and M. M. Ward, "Building a hospital referral expert system with a prediction and optimization-based decision support system algorithm," *Journal of Biomedical Informatics*, vol. 41, no. 2, pp. 371–386, 2008.
- [86] N. Pombo, N. Garcia, and K. Bousson, "Classification techniques on computerized systems to predict and/or to detect apnea: a systematic review," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 265–274, 2017.
- [87] J. Szymański and J. Rzeniewicz, "Identification of category associations using a multilabel classifier," *Expert Systems with Applications*, vol. 61, pp. 327–342, 2016.
- [88] J. Pinho Lucas, S. Segre, and M. N. Moreno, "Making use of associative classifiers in order to alleviate typical drawbacks in recommender systems," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1273–1283, 2012.
- [89] R. Espinosa, D. García-Saiz, M. Zorrilla, J. J. Zubcoff, and J. N. Mazón, "S3mining: a model-driven engineering approach for supporting novice data miners in selecting suitable classifiers," *Computer Standards & Interfaces*, vol. 65, pp. 143–158, 2019.
- [90] D. Cournapeau, "Scikit-learn," 2019.
- [91] N. Hug, "Surprise," 2019.
- [92] M. Kula, "LightFM," in *Proceedings of the 2nd Workshop on New Trends on Content-Based Recommender Systems Co-Located with 9th ACM*, Vienna, Austria, September 2015.
- [93] K. Vand, "Rexy," 2019.
- [94] A. S. Foundation, *PredictionIO*, A. S. Foundation, Pune, Maharashtra, 2019.
- [95] G. Jenson, "HapiGER," 2019.
- [96] L. C. A. and Credits, "LensKit," 2019.
- [97] I. SuggestGrid, "SuggestGrid," 2019.
- [98] S. Systems, "SLI Systems Recommender," 2019.
- [99] A. W. Services, "AmazonWebService Machine Learning," 2019.
- [100] Microsoft, "Azure ML Studio," 2019.
- [101] Gravity Research & Development, "Yusp," 2019.
- [102] IBM Watson Studio, "IBM Watson," 2019.
- [103] Recombee, "Recombee," 2019.
- [104] Mr. Dlib, "Mr. DLib," 2019.
- [105] Caret, "Caret," 2019.
- [106] Shiny, "Shiny," 2019.
- [107] RandomForest, "RandomForest," 2019.
- [108] KLaR, "KLaR," 2019.
- [109] CORElearn, "CORElearn," 2019.
- [110] RecommenderLab, "RecommenderLab," 2019.
- [111] F. J. Domínguez-Mayo, M. J. Escalona, and M. Mejías, "QuEF (quality evaluation framework) for model-driven web methodologies," in *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2010.



Hindawi

Submit your manuscripts at
www.hindawi.com

