CATEGORIZATION OF FACTORS AFFECTING CLASSIFICATION ALGORITHMS SELECTION

Mariam Moustafa Reda, Mohammad Nassef and Akram Salah

Computer Science Department, Faculty of Computers and Information,
Cairo University, Giza, Egypt

ABSTRACT

A lot of classification algorithms are available in the area of data mining for solving the same kind of problem with a little guidance for recommending the most appropriate algorithm to use which gives best results for the dataset at hand. As a way of optimizing the chances of recommending the most appropriate classification algorithm for a dataset, this paper focuses on the different factors considered by data miners and researchers in different studies when selecting the classification algorithms that will yield desired knowledge for the dataset at hand. The paper divided the factors affecting classification algorithms recommendation into business and technical factors. The technical factors proposed are measurable and can be exploited by recommendation software tools.

KEYWORDS

Classification, Algorithm selection, Factors, Meta-learning, Landmarking

1. INTRODUCTION

Due to the large amount of data that is collected nowadays by businesses, an urgent need emerged to process these data and extract some useful information, that can be used in different tasks by the businesses.

Classification, is one data mining task, and can be defined as the process of finding a model that describes and distinguishes data classes. A model to be generated needs a classification algorithm. The classification algorithm learns first on the training data and then generates a model that is ready to score new, unseen data.

There are many popular examples for classification models, one is detecting spam emails. First, spam emails as well as non-spam emails are fed to the algorithm, to generate a model that can be used to later for detection.

More real-life examples of classification problems are: weather forecasting, bank churning and medical diagnosis.

To ensure quality in data mining projects, it is highly recommended to enforce a standard methodology [1]. The most popular methodologies followed by researchers are CRISP-DM: Cross-industry standard process for data mining and SEMMA: Sample, Explore, Modify, Model, and Assess. CRISP-DM was founded by the European Strategic Program on Research in Information Technology, while SEMMA was developed by SAS Institute. Both of these methodologies have well-defined phases for modelling the data by an algorithm and evaluating the model after being created. Also, the first methodology; KDD: Knowledge Discovery in Database was adopted for years by data scientists.

DOI: 10.5121/ijdkp.2019.9401

The Modelling phase is CRISP-DM is equivalent to Model phase in SEMMA and Data Mining in KDD. During this phase, for each data mining task, there are plenty of algorithms that could be used to perform the same data mining task and still produce different results.

In Table. 1, all the phases of the three methodologies mentioned are presented. None of these methodologies defined explicitly a phase for assessing the dataset in hand along with the algorithms available to select the most appropriate algorithm for addressing a data mining task before modelling. This introduces the challenge of selecting the most appropriate algorithm for a data mining task depending on evaluation criteria.

Table 1. Data mining methodologies phases

Knowledge Discovery in Database - KDD	**Sample, Explore, Modify, Model, and Assess - SEMMA**	**Cross-industry standard process for data mining - CRISP-DM**
Pre KDD	-	Business Understanding
Selection	Sample	Data Understanding
Pro processing	Explore	
Transformation	Modify	Data Preparation
Data Mining	Model	Modelling
Interpretation/Evaluation	Assessment	Evaluation
Post KDD	-	Deployment

For example, in the classification task, there are different categories of algorithms each with a list of algorithms; Neural Nets, are one category, where it has a lot of variants and considered as a black box model. Another option is C5.0 algorithm and CHAID algorithm, which are considered as algorithms from the decision trees category. Last but not least, one can consider using a statistical model with all its assumptions about the data.

The overload of choices for the classification task, makes the selection process difficult, in terms of wasted resources. To select the most appropriate classification algorithm to generate the most adequate model to be used is not an easy task.

Practically, it is an iterative work. Given the findings of the algorithms (the models) identified as potential candidates, the models are evaluated and ranked according to criteria such as model accuracy or time complexity. Later on, models with high quality are evaluated according to how the data mining results achieve the business goals.

The selection process can be simple, if there is a list of factors to be considered when selecting the most appropriate classification algorithm for the problem in hand.

According to the no-free-lunch theorem [3], no algorithm is expected to perform better than any other algorithm on all datasets. The assumptions of one good model for one dataset may not hold for another dataset, so it is common in classification task to try multiple algorithms with different configurations and generate multiple models (even more than one for the same algorithm, due to the changes in configuration parameters) and find a model that works best for a particular dataset. Due to this challenge, of finding the most appropriate classification algorithm with the correct configuration parameters for a particular dataset, data miniers and researchers studied and assessed the factors considered in this selection process. The main goal of these studies is offering the guidelines in terms of factors to ease the selection process for experts and non-experts.

The algorithm selection problem was described by Rice, [2]. Considering the algorithm selection problem, in data mining, for the classification task, the portfolio of algorithms will consist of classification algorithms, the instances will be the datasets and the evaluation criteria could be, for example, the accuracy of the model generated by the algorithm. So, the goal is to predict which classification algorithm will have the best accuracy or small error on each dataset.

Multiple systems have been developed since then to tackle the classification algorithm selection problem. These systems perform algorithm selection based on different factors. Depending on these factors one or more algorithms are selected as the most appropriate for the dataset at hand. The system's selection is then justified by evaluating the performance of the selected algorithm(s) compared to the other algorithms depending on some criteria, like classification accuracy or even comprehensibility of the results.

Many approaches followed by data scientists to determine and measure effectiveness of these factors depended on meta-learning. Where data characteristics were extracted and grouped into simple measurements, information theoretic measurements and discriminant analysis measurements. In these studies, the predictors were the data characteristics and the target was the algorithm which performs the best on these data. Other researchers combined algorithms characteristics along with data characteristics to find out the most appropriate algorithms.

Landmarking is another source of dataset characterization. Landmarking concept is to exploit the information obtained on samples of the dataset. The accuracy results from these dataset samples act as the characteristics of the dataset and are referred to as sub-sampling landmarks. These characteristics are then used to guide in the selection of an appropriate classification algorithm for the dataset of interest [4].

2. RELATED WORK

In this section, some of the work and studies done which are related to the factors affecting the problem of selecting the most appropriate classification algorithm for a particular dataset were briefly covered.

As known, each classification algorithm (or category of classification algorithms) has its own basic logic, advantages, disadvantages, configuration parameters and assumptions. Likewise, each dataset has its very own characteristics, which doesn't always fulfill the assumptions of a classification algorithm.

Brute-force is one popular approach in addressing the classification algorithm selection problem. The approach works as follows; iteratively apply available candidates - classification algorithms - on the dataset at hand with fine tuning each of these algorithms' configuration parameters in each iteration. Then, rank the classification algorithms according to some evaluation criteria, that will provide the most suitable results. Later, select the most appropriate algorithm - or algorithms - as the best candidate(s) for classification.

Following the brute force approach would waste a lot of resources. Consequently, researchers study the factors that affect the selection of the appropriate classification algorithm for a dataset and produce tools to recommend the most appropriate classification algorithm for a dataset.
Several studies have proposed the factors and proposed different techniques for dataset characterization to tackle the problem.

[7] proposed a conceptual map of the common knowledge models techniques and intelligent data mining techniques recommender tool based on some dataset characteristics. There was no study carried out to show on which basis were these dataset characteristics used.

In [8], based on the characteristics of datasets and the performance of classification algorithms, mapping between the datasets and the benchmark performance of different classifiers is carried out. K-similar datasets are returned and then ranking of classification algorithms is done so that a classification algorithm is recommended for the dataset at hand.

A subset of the dataset meta-features/characteristics was used without a mention of why this specific subset was favoured over the rest of the available meta-features.

Statlog [10], considered different meta-features in the study and some non-technical factors affecting the classification algorithm selection problem as well. New dataset characteristics extraction approaches like model-based and landmarking weren’t considered.

Although in [18] exhaustive study has been carried out to evaluate the meta-features all together, other non-technical factors weren’t discussed.

[19] proposed Algorithm Selection Tool (AST) based on Case-Based Reasoning. Data Characterization Tool (DCT) developed by Guido Lindner and Robert Engels was used to computes data characteristics. All the dataset characteristics extracted were used as is. No study of the impact of different factors on the selection process was carried out.

[35] carried out a survey for meta-learning with landmarking. The current studies and work related to Landmarking in meta-learning were reviewed and presented. The survey didn’t consider the other approaches used in meta-learning After reviewing the papers in the literature, the following limitations were found. None of the papers considered all the meta-features as well as the business/non-technical factors. The work was done to produce a classification algorithm recommendation tool used only a subset of the dataset meta-features. It was never mentioned explicitly on what basis was this subset selected.

3. METHOD

Basically, if there are sufficient valuable information about the classification algorithms and the datasets available, it would be relatively easy to construct the rules or to build a system that can select the most appropriate classification algorithm for a dataset based on certain characteristics of the problem to solve.

Practically speaking, this is not the usual case except in certain cases. For complex algorithms and systems, there is no full understanding for the factors that affect the performance of a certain classification algorithm on a specific problem that can allow decision making of the classification algorithm selection problem.

As outlined here, a common approach to overcoming these difficulties in selecting the appropriate classification algorithm is to have all the factors affecting this selection listed and categorized and used appropriately to tackle the decision problem.

The related work mentioned the factors considered when selecting the classification algorithms to solve the problem. The analysis of these factors allowed drawing a list of factors that could be considered when selecting the most appropriate classification algorithm, and, classifying these factors into meaningful categories.

Although there were a lot of systems for meta-data extraction and usage were studied and implemented as a system, it was thought that studying the factors that affect the selection making of classification algorithms and categorizing them is more suitable for a practical system used by experienced as well as inexperienced users. Therefore, this paper provides a survey of the different factors considered by researchers, data miners and business experts when selecting the most appropriate algorithm for the dataset at hand. This survey classifies different factors into categories, and shows the importance of each category depending on the related studies.

4. FACTORS CATEGORIZATION

The factors considered when selecting the appropriate classification algorithm as reported in the literature were collected and listed. The goal is to collect all these factors, list them and categorize them. So that, they can be used as a guidance in the classification algorithms selection problem. Figure 1 shows a summarized categorization of the factors listed in this paper.

The literature on classification algorithms and factors affecting the selection of the appropriate algorithm for the dataset at hand was conducted to give insights into how the data miners select an algorithm to use in classification tasks. The authors mentioned the factors used, others mentioned the relevant importance of these factors according to the conducted studies. The analysis of these factors allowed the induction of a categorization tree of factors that can be taken into account when selecting the most appropriate classification algorithm for a dataset.

This paper considers the classification task, and the factors affecting the choice of classification algorithms. Although some factors are common with other data mining tasks, the focus of this paper is the factors affecting algorithm selection for the classification task.

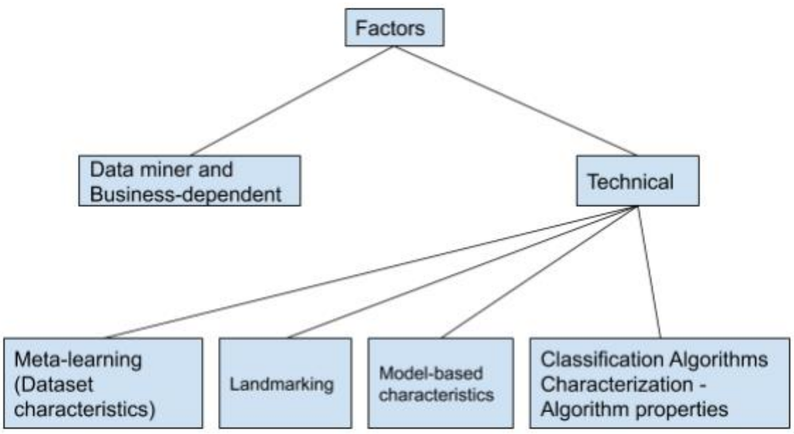

Figure 1. Factors categorization

Table 2. Categorization of factors affecting classification algorithms recommendation

<table>
<tr><td rowspan="3">Data miner and Business-dependent Factors</td><td colspan="2">Data miner’s proficiency in the business domain</td></tr>
<tr><td colspan="2">Familiarity with an algorithm</td></tr>
<tr><td colspan="2">Algorithm’s ease of use and comprehensibility of the results</td></tr>
<tr><td rowspan="4">Technical Factors</td><td rowspan="3">Dataset-dependent</td><td>Meta-learning</td></tr>
<tr><td>Landmarking</td></tr>
<tr><td>Model-based meta-data</td></tr>
<tr><td>Algorithm-dependent</td><td>Characterization of Classification Algorithms</td></tr>
</table>

5. Data Miner and Business-Dependent Factors

These factors are not technical. These are the factors that depend on the business domain, business needs and the data miner experience in the business area of interest where classification is applied, e.g. the data miner performing classification tasks in the banking sector, should have the minimum required knowledge in the banking industry.

5.1. Data miner’s proficiency in the business domain

The proficiency in the business domain - along with familiarity with the business data and systems - plays a critical role in the process of selecting the most appropriate algorithm that can achieve the business objectives. One of the main tasks in a data mining project is translating the business objectives into specific data mining goals. Therefore, the data miner is required to have the business knowledge in the business domain, so that there is consciousness with the possible challenges that could arise ahead in the data mining process.

5.2. Familiarity with an algorithm

On one hand, the expertise of data miner is very valuable, as selecting an algorithm that will solve the classification task is often challenging and requires a lot of research to be done in addition to studying the datasets metadata as well as the algorithm characteristics. On the other hand, prior experience with a certain algorithm may influence the data miner’s decision as it could make him/her biased towards the algorithm he/she is familiar with.

Data miners may embrace the algorithms they are conversant with though there is a possibility that these algorithms may not be the most appropriate for the task to be performed [5].

5.3. Algorithm’s ease of use and comprehensibility of the results

There are two subjective evaluation criteria for a classification algorithm: comprehensibility of the results and the ease-of-use of the algorithm[10]. While the ease-of-use of an algorithm concerns the data miners, the comprehensibility of algorithm’s results mainly concerns’ the stakeholders.

From data miner's view, an algorithm's use is considered easy if its implementation is simple and quick. The ease-of-use of an algorithm is relatively based on data miners experience with the algorithm. Moreover, with regard to the algorithm's configuration parameters, determining their correct values has high computational cost, so an algorithm that has rational defaults or requires fine tuning is considered easy-to-use [10, 26].

Comprehensibility of algorithm's results is as crucial as algorithm's ease of use. Some algorithms can produce more easy to interpret and understand results than others. Depending on the business, an algorithm result must be explained. For example, if-then-else rules are often considered more understandable than a neural net. Neural nets are generally viewed as powerful black box algorithms especially for classification tasks, but the interpretation of the results of the mathematical model that is used behind the scenes is difficult to comprehend than the other types of models [22, 50, 53]. In this case, the business context is the first decider in the selection of the algorithm. A business domain like banking sector might resist using black box algorithms and prefer algorithm with comprehensive results as decision trees, despite the higher accuracy of neural nets. This is mainly due to the ability of decision tree algorithm to generate a ruleset, in the form of if-then-else. These rules can then be revisited by the business experts and stakeholders due to the ease of interpretation of the results. For neural nets, there are no universally accepted guidelines for its use or its complexity. [17, 45].

Consequently, based on results comprehensibility and evaluation criteria, data miners may proceed the classification task with an algorithm that in regards to evaluation criteria is acceptable i.e. optimum, not global optimal, and in terms of results is comprehensive. [5, 31].

Finally, selecting the most appropriate classification model in real life is not only an applied mathematics problem. The algorithm selection process requires: business domain's rules and regulations consciousness, stakeholders' interests consideration and business expertise knowledge involvement. In practice it is essential to combine and balance these factors as this can create the most value.

6. TECHNICAL FACTORS

There are different techniques for tackling the algorithm selection problem. Researchers depend on different technical factors to build an automated system that tackles the problem. These factors are directly related to the dataset characteristics or the classification algorithm parameters or characteristics.

6.1. Meta-learning

The most appropriate algorithm for modelling a particular dataset depends crucially on the metadata of that dataset [10]. Many studies were carried out considering classification algorithms performance with respect to datasets metadata [22, 31, 56]. The fact that dataset metadata and implementation details may influence the accuracy of an algorithm cannot be denied [22]. All dataset metadata examined so far during the implementation of classification recommendation systems were found to affect the success of classification algorithms rate significantly [5, 10].
Meta-learning exploits the datasets' characteristics. Different metadata are presented in meta-learning. These features are divided into several categories [18]. Researchers used meta-learning along with historical performance results of classification algorithms to select the most appropriate algorithm on the current dataset. The term meta-learning stems from the fact that the system tries to learn the function that maps metadata to classification algorithm performance estimates [15]. It is used to gain insight into the algorithm's behaviour with datasets with certain meta-characteristics. Meta-learning might significantly reduce the development time of a

classification task by decreasing the required level of expertise for selecting a suitable classification algorithm for a given dataset [18].

A functional use of meta-learning is building a system that maps an input space consisting of datasets to an output model space consisting of classification algorithms [16]. Different evaluation criteria could be used to evaluate the systems built; mostly accuracy, computational complexity, robustness, scalability, integration, comprehensibility, stability, and interestingness [22]. Several approaches have been developed in this area and it was reported that, regardless of the approach used by the system to select the most appropriate algorithm, the selected algorithm, has high chances of good performance [26].

Meta-learning can address other problems, other than the algorithm selection problem. The authors in [29] attempted to solve a major issue in knowledge data discovery process. The issue of data preprocessing. It goes without saying, that any change in the preprocessing techniques applied on a dataset, can affect the classification algorithm's accuracy and/or speed [22]. The system developed by [29] gives advice for preprocessing. The advice is based on the metadata extracted by the DCT; Data characterization Tool. The research results showed advantageous preprocessings, based on the several test statistics that were calculated. The metadata was used to provide an indication of dataset's complexity.

Although a lot of attention was paid to data preprocessing importance, it can't replace the importance of classification algorithm selection problem. It was reported that the classification algorithm selection process is very important despite the data preprocessing [22]. The nature of dataset determines the most appropriate classification algorithm for it.

Depending on studies of meta-learning, multiple systems have been developed. By using meta-learning, classification algorithms can be accurately recommended as per the given data [8, 22]. Many different metadata have been proposed in the literature. This meta-data is obtained from different concepts, thus can be assorted into three main groups: simple, statistical, and information-theoretic [18]. The central idea is that high-quality metadata provide information to differentiate the performance of a set of classification algorithms [16].

There are different approaches to address the algorithm selection problem. Ofttimes, the selection method is not compared with the other methods [26]. Systems considered so far involved; 1-Case-based reasoning systems, the system has the ability to reason its selection by keeping track of how a problem is solved [25], along with knowledge about past problems. [19] is a case-based reasoning system supporting classification algorithm selection. 2- Classification or regression: algorithm selection task is a classification task that can be solved using a classification or regression algorithm to predict the most appropriate classification algorithm, using a dataset of historical datasets metadata along with classification algorithms portfolios.

For non-experts, it is recommended to use case-based reasoning systems for the algorithm selection problem [26] due to its simplicity. Case-based reasoning algorithms achieve high performance in the algorithm selection task, with respect to the number of historical datasets considered.

Besides manual meta-feature selection, there is work on automatic feature selection for meta-learning [18]. Sometimes reducing the set of metadata increases the performance of a meta-learning system [32].

There is a need to develop an adaptive system which will be smart enough to select the most appropriate classification algorithm [8]. In [25] different approaches to exploit meta-learning to select the appropriate algorithm were discussed. [25] discussed a perspective on how to exploit metadata and build a dynamic learner, that improve their bias dynamically through experience by piling up meta-knowledge. The interesting approach Dynamic-bias selection was discussed. Dynamic-bias is about considering using different subsets of metadata in the dataset meta-learning studies. Since the algorithm selection task is itself a classification task, so different feature selection approaches studies could be applied to metadata as well. In this case, the dataset used is a dataset of metadata about historical datasets, where each dataset is represented in one row, as an example/instance and each attribute represents a metadata measure.

[27] proposed a factor analysis for datasets with a large number of attributes so that the independence among the factors is ensured and the importance level can be measured and new factors can be discovered. Using the method proposed In this study, relevant datasets metadata can be selected based on necessary and sufficient conditions of significance and completeness.
Not only meta-learning that could affect the choice of an algorithm, recent experiments suggested that parameter tuning may affect the classification algorithm accuracy notably [8], but not all of the studies considered parameter tuning. On the same hand, some data preprocessing attempts can affect on the accuracy for some classification algorithms [22]. Keep in mind that the choice of an appropriate feature selection method depends on various dataset metadata; data types, data size and noise [28].

6.1.1. Dataset metadata: Simple, Statistical, Information theoretical

Simple metadata or general data characteristics are measurements which can be easily calculated as they are obtained directly from the data [8]. Statistical metadata is mainly discriminant analysis and other measurements, which can only be computed on continuous attributes. Statistical metadata depicts the statistical properties of the data, e.g. kurtosis. Information theoretical, are metadata which can only be computed on categorical attributes. Although statistical metadata is originally developed for continuous attributes while information theoretical for categorical, both metadata types can be converted to each other, by discretization [18]. A collection of dataset metadata used in different studies is presented in Appendix B.

Statlog project [10], is a comparative study of different classification algorithms. The project attempted to describe datasets in terms of meta-data as a meta-learning step towards creating if-then-else rules that identify under what circumstances which classification algorithm is feasible [6]. These results can be exploited to build models that specify when each algorithm is feasible. The results are strong hardcoded rules or guidelines to guide the algorithm selection process. The metadata considered by Statlog were simple and statistical. Statlog compared the performance of 23 algorithms from symbolic learning, statistics, and neural networks on 21 datasets for the classification task. In StatLog, most of the algorithms had a tuning parameter that was set to its default value, when feasible. Datasets were preprocessed, and the algorithms were evaluated based on the number of criteria. Three of the evaluation criteria were objectively measurable: accuracy, misclassification cost, and the time taken to produce results the other two were subjective: comprehensibility of the results and the ease-of-use of the algorithm to users with relatively little or no experience. As concluded by Statlog, different learning methods are suitable for different problems. The guiding rules concluded by Statlog listed at [10], they were all dependent on the dataset metadata. The ruleset can be turned into a system of if-else and recommend an algorithm for a dataset accordingly.

The Data characteristics tool (DCT), is implemented in a software environment (Clementine) [9]. The DCT is widely used for calculating the three dataset metadata groups about a given data set. In [8], algorithm selection is proposed for classification tasks, by mapping the metadata of datasets extracted by DCT and the performance of classification algorithms. Then for a new dataset, metadata are again extracted using DCT and K-similar datasets are returned. Then ranking of classification algorithms is performed based on performance, and classification algorithm recommended for the problem at hand is based on the highest rank. The study was based on 11 simple metadata, 2 statistical and information theoretical. Results were generated using nine different classification algorithms on thirty-eight benchmark datasets from the UCI repository. The proposed approach used a K-nearest neighbour algorithm for suggesting the most appropriate algorithm. Experimentation showed that predicted accuracies for classification algorithms are matching with the actual accuracies for more than 90% of the benchmark datasets used. It was concluded that the number of attributes, the number of instances, number of classes, maximum probability of class and class entropy are the main metadata which affects the accuracy of the classification algorithm and the automatic selection process of it.

Another large-scale project that utilizes meta-learning is METAL [11]. METAL's main objective was enhancing the use of data mining tools and specifically to expand savings in the experimentation time [16]. The METAL project [13] focused on finding new and significant data characteristics. It used metadata of the datasets along with the classification algorithms to learn how they can be combined. The project resulted in the Data Mining Advisor (DMA) [12]. DMA is a web-enabled solution that supports users in algorithm selection by automatically selecting the most appropriate classification algorithms. It was developed as an implementation of a meta-learning approach. DMA provides recommendations for classification algorithms in the form of rankings. A list ordered from best to worst is produced. The list is sorted in consonance with a weighted combination of parameters as accuracy and time taken in training [16]. DMA uses the DCT and a k-Nearest Neighbor algorithm to rank ten target classifiers. The DMA presented two different evaluation approaches for the ranking of the classification algorithms; first technique makes use of the ratio of accuracy and training time and the other ranking technique is based on the concept of data envelopment analysis [14].

[18] performed an exhaustive evaluation of the three dataset metadata categories along with other factors for meta-learning using regression. The research was based on 54 datasets from the UCI machine learning repository and from StatLib. It was concluded that utilizing the dataset metadata for algorithm selection performs better than the baseline. [18] utilized the Pearson product-moment correlation coefficients, to automatically select highly correlated metadata from the metadata groups for the set of target classification algorithms. It was shown that the automatic feature selection selects the most useful metadata. This is one recent research area, utilizing automatic feature selection techniques to select the most significant metadata measures.

DM assistant tool [7] and Algorithm Selection Tool, AST [19] use a case-based reasoning approach to support classification algorithm selection. AST benefits from data characteristics extracted by DCT and considered application restrictions for the algorithm selection process. AST gives the user recommendation which algorithm should be applied, along with an explanation for the recommendation in the form of past experiences available in the case base. A new algorithm can be added to the case base of AST easily without testing on all historical datasets. [19] considered the use of all of the three dataset metadata categories in building AST. The metadata was used to compute the most similar cases. All the classification algorithms of the case base were tested with their default parameters values, no fine tuning for the parameters. Also, AST had no preferences in the metadata extracted by DCT, they were all used with equal importance. The results were evaluated, overall the accuracy of ACT for the most appropriate algorithm of the first similar case is applicable in 79%. For datasets with only continuous attributes or with continuous and categorical attributes, the rate is over 85%. While datasets with only categorical attributes are less than 68%. This is an indicator that the metadata for the categorical attributes are still insufficient and those additional measurements are required. Fine tuning for the categorical attributes or selecting the most relevant ones could enhance the accuracy of AST for datasets with only categorical attributes. Some of the metadata may be irrelevant, others may not be adequately represented, while some important ones may be missing [24].

DM assistant offers the users the most appropriate data mining techniques for the problem of interest. The system automatically extracts the most relevant metadata from a given dataset to find the most similar cases. Some of the metadata extracted by DM assistant are the number of classes, the entropy of the classes and the percent of the class mode category [7]. There were no specific details of the metadata extracted and used to measure the distance with historical datasets to find the most relevant.

In [21] the complexity of each dataset was measured by considering its metadata. The three metadata categories; simple, statistical, and information theoretic were considered for each dataset. A total of 21 metadata measures were calculated. Most of the metadata measures described in [23]. The overall accuracy of the system in predicting the most appropriate classification algorithm is 77%. This makes confidence in the metadata measures used by [21] good enough to be used in other studies. [21] trained a neural network to predict a classification algorithm performance. Dataset metadata are fed as input to the neural network, and the output is a ranked list of techniques predicting their likely performance on the dataset. To model the classification algorithms performance, 57 datasets were used from the UCI machine learning repository, a total of 21 metadata measures that describe the characteristics of the data were calculated. And six classification algorithms were modelled.

The goal of [20] was to assist users in the process of selecting an appropriate classification algorithm without testing the huge array of classification algorithms available. [20] aimed to determine the dataset metadata that lends themselves to superior modelling by certain classification algorithms by introducing a rule-based classification approach (C5.0) for classification algorithm selection. Most of the generated rules are generated with a high confidence rating. The metadata of the datasets used in this study described in [21]. The metadata of each dataset was extracted and quantitatively measured, it was combined along with the empirical evaluation of classification algorithms performance, to generate the rules. The rules generated for all 8 classification algorithms based on the classification performance of 100 datasets.

[24] presented a meta-learning approach to boost the process of selecting an appropriate classification algorithm. It used the k-Nearest Neighbor algorithm to detect the datasets that are closest to the dataset of interest. The research studied the importance of a relatively small set of dataset metadata, but it is believed that this small set of metadata provide information about properties that affect algorithm performance. Performance of the candidate classification algorithms on the datasets was used to recommend the most appropriate algorithm to the user in the form of ranking. The algorithm’s performance is evaluated using a multicriteria evaluation measure that considers the accuracy and the training time. Results show, most of the metadata used were useful to select classification algorithms on the basis of accuracy. To avoid bias, the author recommended using feature selection methods at the meta-level to select the appropriate metadata for a given multicriteria setting. A visual analysis of the set of metadata was performed aiming to identify the measures that appear to provide less useful information. The visual analysis was done by analyzing the correlation between the values of a specific meta-attribute and each algorithm’s performance. The research metadata used were simple, statistical and information-theoretical, described in detail by [23].

[22] used metadata that represents a set of characteristics that affect the classification algorithms’ performance. Regression models developed in this study that offer hints to data miners about the classification algorithm expected accuracy and speed based on dataset metadata. Moreover [22] studied the correlations between dataset metadata and accuracy and it was found that all these metadata can affect the classification algorithm performance, i.e. make a significant difference in the classification algorithms success rate. Criteria used in the classifiers evaluation are mostly

accuracy, computational complexity, robustness, scalability, integration, comprehensibility, stability, and interestingness. Ten datasets collected from the UCI Machine Learning Repository were used to run the 14 classification algorithms. Datasets were preprocessed and all numeric attributes in the datasets were converted to categorical attributes by binning them into intervals within ±1 standard deviation and saved as new attributes. The results showed that some of the classification algorithms studied cannot handle continuous variables and dense dimensionality. Moreover [22] claimed that the metadata: the high number of variables and the high number of instances increase the classification task difficulty and impact the algorithm classification power. In summary, all dataset metadata were found to affect the success rate significantly.

[30] exploited DCT to extract metadata from datasets. The three metadata categories were considered. [30] proposed a Zoomed ranking technique. In the zooming phase, the k-Nearest Neighbor algorithm is employed with a distance function based on a set of dataset metadata to identify datasets from previously processed datasets, that are similar to the one at hand. These datasets performance information is expected to be relevant for the dataset at hand. In ranking phase, the adjusted ratio of ratios ranking method is used. The ranking is on the basis of the performance information (accuracy and total execution time) of the candidate algorithms on the datasets selected in zooming phase. [30] made no investigation on the metadata used, whether they are relevant or not. And, no investigation was made to determine if different weights should be assigned to them in the distance function. Metadata measures were chosen because they are provided by DCT and were used before for the same purpose. Although no statistical support, it was claimed that zooming improves the quality of the rankings generated, which gives an indication that the metadata used in the study is good enough to be used in other studies.

Although all of the studies discussed here made heavy use of metadata of datasets, and showed the different techniques necessary to build effective meta-learning systems, it is emphasized the importance of studying alternative meta-features in the characterization of datasets [16].

There are a lot of studies for the metadata extracted from the datasets. This unleashes two research questions, 1- should different dataset metadata be considered? 2- How good are the available feature selection techniques in selecting significant metadata.

6.2. Landmarking

Landmarking is a new and promising approach to extract metadata [35], that utilizes simple and fast computable classification algorithms [18]. Landmarking attempts to determine the position of a specific dataset in the space of all available historical datasets by directly measuring the performance of some simple and significant classification algorithms themselves [25, 34, 35]. One idea of landmarking is about characterizing datasets by classification algorithms themselves. Landmarking features can be seen as dataset characteristics, where these characteristics represent the performance of some fast, simplified versions of classification algorithms on this dataset [15, 24]. These simplified versions of the algorithms are called landmarks [37]. This means that landmarks are estimates to the performance of the full version of an algorithm for a given dataset. There are some conditions that have to be satisfied when choosing a landmark, described in [37]. Based on some classification algorithm evaluation criteria, one algorithm is the winner over another for the dataset at hand.

Another idea of landmarking is to exploit information obtained on samples of datasets and the full version of the algorithm. Accuracy results on these samples serve to characterise the datasets and are referred to as sub-sampling landmarks. This information is subsequently used to select the most appropriate classification algorithm [16].

Experiments showed that landmarking selects with a mild and rational level of success, the best performing algorithm from a set of classification algorithms [34]. Experiments show that landmarking approach compares favourably with other meta-learning approaches [38]. It was reported that landmarking-features are well suited for meta-learning [37]. The landmarking features are acceptable and can be used to build systems for selecting the most appropriate classification algorithm for a dataset. Using landmarking features for predicting the most appropriate classification algorithm - out of pair and out of all available classification algorithms - has been evaluated by different researchers [18].

Although many research studies were done in the area of landmarking there are still many open challenges in this area that need additional experiments and research.

6.3. Model-based Metadata

Model-based metadata is a decision tree model - without pruning - different properties, created from the dataset [18,35]. Examples of decision tree model properties are number of leaves, number of nodes, nodes per attribute, nodes per sample and leaf correlation [35]. The hypothesis is that the decision tree model induced from datasets owns the characteristics that are highly dependent upon the dataset. There is a number of important connections between dataset characteristics and induced trees, the properties of the induced tree model are mapped to data characteristics in [39].

In this approach, instead of using classification algorithms' performances to describe datasets, as in landmarking, or metadata measures of datasets, as in the traditional approach algorithm's hypotheses were used [39].

6.4. Characterization of Classification Algorithms

There are many classification algorithms available in the literature, all of them need one or more attribute to use as predictor to divide the data, and a target to predict. Classification algorithms are grouped according to their characteristics. Many studies were conducted to compare the classification algorithms in terms of performance: Accuracy, Complexity, and Training Time [22, 31, 43, 45, 47, 56]. Studies revealed how the feature selection could improve the classification ability of classification algorithms [28, 44]. Studies considered combination/ensemble of the models of several algorithms as it usually has a positive effect on the overall quality of predictions, in terms of accuracy, generalisability and/or lower misclassification costs [42, 44, 53] For instance, random forest, is ensemble algorithm, and it is known as one of the accurate classification algorithms [43].

Although algorithms within the same group, share many characteristics like how new instances are scored, differ in the other characteristics. Each group has its strengths and weaknesses [10, 42, 43]. The classification algorithms were mainly grouped into three different groups [10, 31]; symbolic algorithms, statistical algorithms and neural nets. A brief description of the categories of classification algorithms is presented in Appendix A.

6.4.1. Predictors fields and target field(s) types

Some classification algorithms can be used with categorical and continuous predictors while others can be used with categorical only [41]. Also, some algorithms are more appropriate than others when it comes to the predicted field, or the target, as some are able to classify only categorical targets. Neural networks, for example, can handle any predictor and target. So, this is a decision making factor, in the process of algorithm selection problem.

A well-known preprocessing task, is binning continuous attributes to convert it to a categorical attribute. This can be used as a tweak to feed continuous data to an algorithm that only works with categorical data.

6.4.2. How are missing data handled

Missing values are a common occurrence in datasets. To handle the missing values in data a strategy needs to be followed, as a missing value can have different meanings in the dataset. Not all classification algorithms treat missing values similarly. Some algorithms can't process missing values in the dataset while others can. Usually, the data miner follows one of these strategies to handle missing data: ignore the missing values, or discard any raw in the dataset containing missing values, or substitute the missing values with the mean if the attribute is continuous, or if the attribute is categorical, either deduce missing values from existing values or consider it as a new value of the attribute. Algorithms handle missing data differently, different ways of handling missing data described in [54].

Handling missed data is a crucial subtask in data preprocessing phase [10, 31, 41, 55]. That is one reason why the ratio of missing values is always importantly considered as a significant statistical metadata measure of the dataset and considered in many pieces of research [19, 20, 21, 24, 30, 37].

6.4.3. Model Assumptions

Each classification algorithm has one or more assumption [52]. One assumption is the normality assumption [10, 29, 52]. Another example is the sample size for neural nets [10, 17, 52]. Linearity between dependent and independent variables and the multivariate normal distribution of the independent variable are other assumptions by different classification groups [10, 52]. The performance of a classification algorithm compared to other candidates depends on the dataset characteristics and how well these characteristics comply with the assumptions made by the classification algorithm [18].

7. CONCLUSIONS

The study of different factors considered when selecting the most appropriate classification algorithm for a dataset showed that the resulted model is sensitive to changes in data characteristics and classification algorithm characteristics. Considering the proposed factors helps data miners recommend an appropriate classification algorithm and build classification algorithms recommendation systems. Generally, more than one factor should be considered to recommend the most appropriate classification algorithm for a dataset.

Basically, the factors listed and categorized in this study can be used to shortlist a few classification algorithms. Selecting the most appropriate classification algorithm to a real life problem requires awareness of business domain's rules and regulations, and consideration of business expertise knowledge.

It was shown that different classification algorithm recommendation systems considered meta-learning, where metadata of the dataset is extracted and studied so that the recommendation system can use the extracted metadata to select the most appropriate algorithm [8, 9, 11, 12, 22]. It was also shown that these metadata can be categorized into simple, statistical and information theoretic [18].

Due to the importance of the stage of selecting the most appropriate classification algorithm in data mining - as it defines the type of results that will be produced, which will later influence the subsequent decisions- different paths were considered to facilitate the classification algorithm recommendation process. It was shown here, landmarking and model-based metadata.

Landmarking exploits simple and fast computable classification algorithms to determine the position of a specific dataset in the space of all available historical datasets by directly measuring the performance of some simple and significant classification algorithms themselves. On the other hand, the model-based metadata utilizes decision tree model - without pruning - different properties. The hypothesis is that the decision tree model induced from datasets owns the characteristics that are highly dependent upon the dataset.

Experiments showed that landmarking approach compares favourably with other meta-learning approaches, but there are many open challenges for further research in the area of landmarking and model-based metadata.

It was also emphasized the characteristics of classification algorithms that can be measured and used as factors to recommend the most appropriate classification algorithm for a dataset.

The main conclusion is that: there is no single factor or group of factors that can be used alone to recommend a classification algorithm for a dataset. Although most of the studies for studying these factors depends crucially on metadata of the datasets. It was shown that there are other paths that can be considered as well in recommending a classification algorithm for a dataset.

In future work, the factors used to recommend the most appropriate classification algorithm for the dataset at hand have to be refined. The main point is to prioritize the metadata extracted for meta-learning, according to their significance.

APPENDIX A: CLASSIFICATION ALGORITHMS CATEGORIES

A.1. Symbolic algorithms

They produce a set of rules that divide the inputs into smaller decisions in relation to the target. Decision trees algorithms belong to this group as the resulting trees can be turned into a set of if-then-else rules easily. Symbolic algorithms are simple, easy to implement, interpret, and represent a good compromise between simplicity and complexity [40, 43, 46]. A lot of studies carried out to describe, review and compare these algorithms [10, 31, 41, 42, 45]. On one hand, it was reported that symbolic algorithms are a good choice for maximizing classification accuracy if the metadata of the dataset shows that the data has extreme distribution [10, 45]. On the other hand, they become poor choice if misclassification cost should be minimized, or the dataset has equally important continuous attributes [10]. Studies were also conducted to compare among available symbolic algorithms and evaluate its performance, in terms of the tree size and complexity and training time, for instance [31].

A.2. Statistical algorithms

The resulted model of a statistical algorithm is expressed by an equation, and statistical tests can lead field selection in the model. Statistical algorithms assume certain distributions in the data, which makes them more harder than symbolic algorithms but still less difficult than neural networks. Because using statistics science with data allows analysis and interpretation of the data, there are many classification algorithms based on statistics and tremendous quantities of research for statistical algorithms [51].

A.3. Neural networks

Neural networks consider the human brain as their modelling tool [22]. Neural networks are used to perform nonlinear statistical modelling, as they have the ability to detect complex nonlinear relationships between dependent and independent attributes in the data [50]. There are multiple training algorithms for neural networks that have been studied and presented, showing how they work [17, 25, 48, 49] They don't produce rules or equations.

Lately, a lot of emphasis has been placed on neural networks, because of it powerfulness in prediction [44, 47, 50]. The accuracy of neural network classifications was found to be significantly influenced by the size of the training set, discriminating variables and the nature of the testing set used [17, 50]. Although its reported robustness, high adaptability and accuracy, neural networks algorithms require large machine resources [10]. Also, neural networks are prone to overfitting. With overfitting, the error on the training set is driven to a very small value, thus the model won't be able to generalize well to new data [5].

APPENDIX B: COLLECTION OF DATASET METADATA USED IN DIFFERENT STUDIES

Simple meta-features: number of samples, number of classes, number of attributes, number of nominal attributes, number of continuous attributes, the ratio of nominal attributes, the ratio of continuous attributes, dimensionality (number of attributes divided by the number of samples).
Statistical meta-features: kurtosis, skewness, canonical discriminant correlation (cancor1), first normalized eigenvalues of the canonical discriminant matrix (fract1), absolute correlation.

Information-theoretic meta-features: normalized class entropy, normalized attribute entropy, joint entropy, mutual information, noise-signal-ratio, the equivalent number of attributes.

REFERENCES

[1] Shafique, Umair & Qaiser, Haseeb. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research. 12. 2351-8014.

[2] Rice, J. R. (1976). The algorithm selection problem. Advances in Computers. 15. 65–118.

[3] Wolpert, David & Macready, William G. (1997). No free lunch theorems for optimization. IEEE Transac. Evolutionary Computation. 1. 67-82.

[4] Soares, C. & Petrak, J. & Brazdil, P. (2001) Sampling-Based Relative Landmarks: Systematically Test-Driving Algorithms before Choosing. In: Brazdil P., Jorge A. (eds) Progress in Artificial Intelligence. EPIA 2001. Lecture Notes in Computer Science. 2258.

[5] Chikohora, Teressa. (2014). A Study Of The Factors Considered When Choosing An Appropriate Data Mining Algorithm. International Journal of Soft Computing and Engineering. 4. 42-45.

[6] Michie, D. & Spiegelhalter, D.J. & Taylor, C.C. (1994). Machine Learning, Neural and Statistical Classification, Ellis Horwood, New York.

[7] Gibert, Karina & Sànchez-Marrè, Miquel & Codina, Víctor. (2018). Choosing the Right Data Mining Technique: Classification of Methods and Intelligent Recommendation.

[8] N. Pise & P. Kulkarni. (2016). Algorithm selection for classification problems. SAI Computing Conference (SAI). 203-211.

[9] Peng, Yonghong & Flach, Peter & Soares, Carlos & Brazdil, Pavel. (2002). Improved Dataset Characterisation for Meta-learning. Discovery Science Lecture Notes in Computer Science. 2534. 141-152.

[10] King, R. D. & Feng, C. & Sutherland, A. (1995). StatLog: comparison of classification algorithms on large real-world problems. Applied Artificial Intelligence. 9. 289-333.

[11] Esprit project Metal. (1999-2002). A Meta-Learning Assistant for Providing User Support in Machine Learning and Data Mining. Http://www.ofai.at/research/impml/metal/.

[12] Giraud-Carrier, C. (2005). The data mining advisor: meta-learning at the service of practitioners. Machine Learning and Applications. 4. 7.

[13] Giraud-Carrier, Christophe. (2008). Meta-learning tutorial. Technical report, Brigham Young University.

[14] Paterson, Iain & Keller, Jorg. (2000). Evaluation of Machine-Learning Algorithm Ranking Advisors.

[15] Leite, R. & Brazdil, P. & Vanschoren, J. (2012). Selecting Classification Algorithms with Active Testing. Machine Learning and Data Mining in Pattern Recognition. 7376. 117-131.

[16] Vilalta, Ricardo & Giraud-Carrier, Christophe & Brazdil, Pavel & Soares, Carlos. (2004). Using Meta-Learning to Support Data Mining. International Journal of Computer Science & Applications. 1.

[17] Foody, G. M. & Arora, M. K. (1997). An evaluation of some factors affecting the accuracy of classification by an artificial neural network. International Journal of Remote Sensing. 18. 799-810.

[18] Reif, M. & Shafait, F. & Goldstein, M. & Breuel, T.M.& Dengel, A. (2012). Automatic classifier selection for non-experts. Pattern Analysis and Applications. 17. 83-96.

[19] Lindner, Guido & Ag, Daimlerchrysler & Studer, Rudi. (1999). AST: Support for algorithm selection with a CBR approach. Lecture Notes in Computer Science. 1704.

[20] Ali, Shawkat & Smith-Miles, Kate. (2006). On learning algorithm selection for classification. Applied Soft Computing. 6. 119-138.

[21] Smith, K.A. & Woo, F. & Ciesielski, V. & Ibrahim, R. (2001). Modelling the relationship between problem characteristics and data mining algorithm performance using neural networks. Smart Engineering System Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Data Mining, and Complex Systems. 11. 357–362.

[22] Dogan, N. & Tanrikulu, Z. (2013). A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness. Information Technology Management. 14. 105-124.

[23] Henery, R. J. Methods for Comparison. Machine Learning, Neural and Statistical Classification, Ellis Horwood Limited, Chapter 7, 1994.

[24] Brazdil, P.B.& Soares, C. & da Costa. (2003). Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results. J.P. Machine Learning. 50. 251-277.

[25] Vilalta, R. & Drissi, Y. (2002). A Perspective View and Survey of Meta-Learning. Artificial Intelligence Review. 18. 77-95.

[26] Kotthoff, Lars & Gent, Ian P. & Miguel, Ian. (2012). An Evaluation of Machine Learning in Algorithm Selection for Search Problems. AI Communications - The Symposium on Combinatorial Search. 25. 257-270.

[27] WANG, HSlAO-FAN & KUO, CHINC-YI. (2004). Factor Analysis in Data Mining. Computers and Mathematics with Applications. 48. 1765-1778.

[28] Dash, M. & Liu, H. (1997). Feature Selection for Classification. Intelligent Data Analysis. 1. 131-156.

[29] Engels, Robert & Theusinger, Christiane. (1998). Using a Data Metric for Preprocessing Advice for Data Mining Applications. European Conference on Artificial Intelligence. 430-434.

[30] Soares C. & Brazdil P.B. (2000). Zoomed Ranking: Selection of Classification Algorithms Based on Relevant Performance Information. Principles of Data Mining and Knowledge Discovery. 1910.

[31] Lim, TS. & Loh, WY. & Shih, YS. (2000). A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms. Machine Learning. 40. 203-228.

[32] Todorovski, L. & Brazdil, P. & Soares, C. (2000). Report on the experiments with feature selection in meta-level learning. Data Mining, Decision Support, Meta-Learning and ILP. 27–39.

[33] Kalousis, A. & Hilario, M. (2001). Feature selection for meta-learning. Advances in Knowledge Discovery and Data Mining. 2035. 222–233.

[34] Pfahringer, Bernhard & Bensusan, Hilan & Giraud-Carrier, Christophe. (2000). Meta-learning by Landmarking Various Learning Algorithms. International Conference on Machine Learning. 7. 743-750.

[35] Balte, A., & Pise, N.N. (2014). Meta-Learning With Landmarking: A Survey. International Journal of Computer Applications. 105. 47-51.

[37] Daniel Abdelmessih, Sarah & Shafait, Faisal & Reif, Matthias & Goldstein, Markus. (2010). Landmarking for Meta-Learning using RapidMiner. RapidMiner Community Meeting and Conference.

[38] Bensusan H., Giraud-Carrier C. (2000). Discovering Task Neighbourhoods through Landmark Learning Performances. Principles of Data Mining and Knowledge Discovery. 1910. 325-330.

[39] Bensusan, Hilan & Giraud-Carrier, Christophe & J. Kennedy, Claire. (2000). A Higher-order Approach to Meta-learning. Meta-Learning: Building Automatic Advice Strategies for Model Selection and Method Combination. 109-117.

[40] Podgorelec, Vili & Kokol, Peter & Stiglic, Bruno & Rozman, Ivan. (2002). Decision Trees: An Overview and Their Use in Medicine. Journal of medical systems. 26. 445-63.

[41] M. AlMana, Amal & Aksoy, Mehmet. (2014). An Overview of Inductive Learning Algorithms. International Journal of Computer Applications. 88. 20-28.

[42] Behera, Rabi & Das, Kajaree. (2017). A Survey on Machine Learning: Concept, Algorithms and Applications. International Journal of Innovative Research in Computer and Communication Engineering. 2. 1301-1309.

[43] Ponmani, S. & Samuel, Roxanna & VidhuPriya, P. (2017). Classification Algorithms in Data Mining – A Survey. International Journal of Advanced Research in Computer Engineering & Technology. 6.

[44] Aggarwal C.C., Zhai C. (2012). A Survey of Text Classification Algorithms. Mining Text Data. 163-222.

[45] Bhuvana, I. & Yamini, C. (2015). Survey on Classification Algorithms for Data Mining:(Comparison and Evaluation). Journal of Advance Research in Science and Engineering. 4. 124-134.

[46] Mathur, Robin & Rathee, Anju. (2013). Survey on Decision Tree classification algorithms for the Evaluation of Student Performance.

[47] Abd AL-Nabi, Delveen Luqman & Ahmed, Shereen Shukri. (2013). Survey on Classification Algorithms for Data Mining:(Comparison and Evaluation). Computer Engineering and Intelligent Systems. 4. 18-24.

[48] Schmidhuber, Juergen. (2014). Deep Learning in Neural Networks: An Overview. Neural Networks. 61.

[49] Kotsiantis, Sotiris. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica (Ljubljana). 31.

[50] Tu, Jack V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of Clinical Epidemiology. 49. 1225-1231.

[51] Bousquet O., Boucheron S., Lugosi G. (2004). Introduction to Statistical Learning Theory. Advanced Lectures on Machine Learning. 3176. 169-207.

[52] Kiang, Melody Y. (2003). A comparative assessment of classification methods. Decision Support Systems. 35. 441-454.

[53] Gama, João & Brazdil, Pavel. (2000). Characterization of Classification Algorithms.

[54] P. Nancy & R. Geetha Ramani. (2011). A Comparison on Performance of Data Mining Algorithms in Classification of Social Network Data. International Journal of Computer Applications. 32. 47-54.

[55] Soroush Rohanizadeh, Seyyed & Moghadam, M. (2018). A Proposed Data Mining Methodology and its Application to Industrial Procedures.

[56] Brazdil P. & Gama J. & Henery B. (1994). Characterizing the applicability of classification algorithms using meta-level learning. Machine Learning. 784. 83-102.

AUTHORS

Mariam Moustafa Reda received B.Sc. (2011) from Fayoum University, Cairo, Egypt in Computer Science. In 2012, she joined IBM Egypt as Application Developer. Mariam has 2 published patents. Since 2014, she started working in data analytics and classification related projects. Her research interests include data mining methodologies improvement and automation.

Mohammad Nassef was graduated in 2003 from Faculty of Computers and Information, Cairo University. He has got his M.Sc. degree in 2007, and his PhD in 2014 from the same University. Currently, Dr Nassef is an Assistant Professor at the Department of Computer Science, Faculty of Computers and Information, Cairo University, Egypt. He is interested in the research areas of Bioinformatics, Machine Learning and Parallel Computing.

Akram Salah graduated from mechanical engineering and worked in computer programming for 7 years before he got his M.Sc. (85) and PhD degrees from the University of Alabama at Birmingham, the USA in 1986 in computer and information sciences. He taught at the American University in Cairo, Michigan State University, Cairo University, before he joined North Dakota State University where he designed and started a graduate program that offers PhD and M.Sc. in software engineering. Dr Salah's research interest is in data knowledge and software engineering. He has over 100 published papers. Currently, he is a professor in the Faculty of Computer and Information, Cairo University. His current research is in knowledge engineering, ontology, semantics, and semantic web.