**RESEARCH PAPER**

# Machine learning in information systems - a bibliographic review and open research issues

**Benjamin M. Abdel-Karim[1]** (ORCID) · **Nicolas Pfeuffer[1]** · **Oliver Hinz[1]**

## Abstract

Artificial Intelligence (AI) and Machine Learning (ML) are currently hot topics in industry and business practice, while management-oriented research disciplines seem reluctant to adopt these sophisticated data analytics methods as research instruments. Even the Information Systems (IS) discipline with its close connections to Computer Science seems to be conservative when conducting empirical research endeavors. To assess the magnitude of the problem and to understand its causes, we conducted a bibliographic review on publications in high-level IS journals. We reviewed 1,838 articles that matched corresponding keyword-queries in journals from the AIS senior scholar basket, Electronic Markets and Decision Support Systems (Ranked B). In addition, we conducted a survey among IS researchers (N=110). Based on the findings from our sample we evaluate different potential causes that could explain why ML methods are rather underrepresented in top-tier journals and discuss how the IS discipline could successfully incorporate ML methods in research undertakings.

## Introduction

The constant evolution of Machine Learning (ML) approaches over the past 20 years has led both practice and research to breakthroughs in technological developments in various areas (Jordan and Mitchell 2015, p. 255). Since the generation and mining of data has become more feasible and large amounts of computing power have become considerably more accessible and affordable in the past decade,

ML methods, with their ability to automatically solve problems with large sets of parameters, have increasingly been applied in many areas (Delen et al. 2013, p. 1152).

In recent years, Deep Learning (DL) methods, a subset of ML methods, have gained greater popularity, especially due to the availability of large amounts of complex data (Agarwal and Dhar 2014, p. 444).

It is undisputed that Computer Science will primarily drive the technical progress of these technologies. The industry also seems to be very interested in the advancement of these methods, as shown by projects like Alpha Go from Google (Silver et al. 2017). However, ML methods are far more than just IT-specific tools and could be useful for decision support, knowledge inference, and process automation for many different industry areas. ML is on its way to becoming a key technology in the digital transformation aimed at boosting productivity and fostering new discoveries across many industries (Jain et al. 2018, p. 250). To achieve a higher level of innovation and to stay competitive in the market, traditional industrial actors may need to enable their employees to apply more profound analytical methods, especially regarding developments in the fields of Artificial Intelligence (AI) and ML (Aleksander 2017).

✉ Benjamin M. Abdel-Karim
Abdel-Karim@wiwi.uni-frankfurt.de

Nicolas Pfeuffer
pfeuffer@wiwi.uni-frankfurt.de

Oliver Hinz
hinz@wiwi.uni-frankfurt.de

[1] Goethe University Frankfurt, Information Systems and Information Management, Theodor-W.-Adorno-Platz 4, 60323, Frankfurt, Germany

However, without theoretical guidelines on how to apply different ML methods on more complex and industry-specific socio-technical problems, whether the goal is the inference of valuable knowledge for strategic decisions or the optimization of business processes, firms from industries beyond IT will constantly be lagging behind large IT firms who find themselves facing an imminent monopoly.

Regarding the research on ML methods, Computer Science fulfills its role in pioneering the development of engineering methods and fundamental research in the field of algorithms. In this regard, Information Systems (IS) research should see its role in the transfer of theoretical knowledge from machine learning to applications that solve (industry-)specific problems, with the goal of advancing IS theory as well as help practitioners in applying ML methods, "to best manage and support IT or IT-enabled business initiatives." (Benbasat and Zmud 2003, p.192). While Computer Science and engineering greatly engaged with this technology's advancement, it seems that IS researchers do not entirely embrace the new possibilities.

More than 30 years ago, Straub (1989) posed the question of whether IS research reflects upon itself critically. As technology is shifting more in the direction of automated and autonomous systems and digital transformation is permeating every industry, IS research should consider paying the necessary level of attention to current technological developments (Boudreau et al. 2001). In particular, the field of ML must not be excluded from consideration during this process (Maass et al. 2018). In recent years, senior IS scholars have pondered the integration of stronger analytical methods and whether to greet them with skepticism or embrace them with open arms. For example, several scholars have been calling for stronger yet rigorous integration of Big Data and connected analytical methods, such as ML, in IS research (e.g., Abbasi et al. 2016; Maass et al. 2018), and have encouraged the field to "embrace the challenges and claim our territory" (Goes 2014, p. viii). While there has been a lot of discussion about data and ML, it remains unclear to what extent the IS discipline has been adapting to the emerging technological phenomenon of ML.

An important point, which is regularly emphasized by senior scholars, is that IS research has to deal with and analyze the IT artifact in particular because, especially in this way, business and society can benefit from our research (e.g., Benbasat and Zmud 2003). Since ML methods offer many possibilities for data aggregation, processing, and analysis, it also seems to provide excellent opportunities for application in research projects. This is especially the case for research on IS-relevant phenomena, which are either data-intensive or related to research on AI-related topics or both (e.g., Abbasi et al., 2016; Maass et al., 2018; ). Even as a non-analytical instrument, ML can be used to build highly functional systems in experimental studies or case studies (e.g., Benbasat et al. 1987), which could significantly increase the validity of such studies. In this context, it is important to realize that ML cannot be a universal tool for all purposes, although the limits of what is possible with ML do not seem to have been sufficiently tested either.

To put it in a nutshell, the expanded use of ML as a research instrument has the potential to enrich research projects methodologically. Furthermore, it may also increase the relevance of the resulting findings and at the same time can help fulfill the mission of IS research with regard to practice: to provide guidance and justify both the application and applicability to ensure "that what is being discovered and applied is in fact 'correct' "(Benbasat and Zmud 1999, p. 12).

In the context of self-reflection of the IS discipline, we, therefore, aim to examine the current state of research in IS for ML and try to answer the following questions in the context of IS research:

RQ 1.1)   How prevalent is the application of ML as a research instrument in IS research?

RQ 1.2)   If the analysis reveals a low number of ML-based papers in top-tier research journals, what may be possible causes for this observation?

This work aims to understand the context of ML (non-) adoption and application as a cutting-edge technology within scientific research in the IS field. Beyond that, this paper also aims to make IS research more aware of the topic of applying ML methods based on our bibliographic review and subsequent scientometric analyses (Leydesdorff 2001; Khan and Wood 2016). By answering our aforementioned research questions, we aim to contribute to IS research in several ways: (1) By analyzing the apparent issue with the application of ML in IS research through a bibliographic bibliographic review in a quantitative manner, we emphasize that a real issue persists. (2) By conducting a nationwide survey involving 110 IS scholars in Germany, we closely examine the core reasons for the issue and explain how these components deter the application of ML methods in IS research. (3) We conclude this article by suggesting implications and a structured research framework to amalgamate the classical theory-driven approach with the more data-oriented ML approach in IS research. By working out key research questions for future work, we finally aim to provide guidelines for IS research to tackle the analyzed problem in a goal-oriented manner.

The remainder of this article proceeds as follows: "Theoretical background" provides a theoretical background of ML and a high-level description of how it is applied. "Methodology" describes our methodology. Following that,
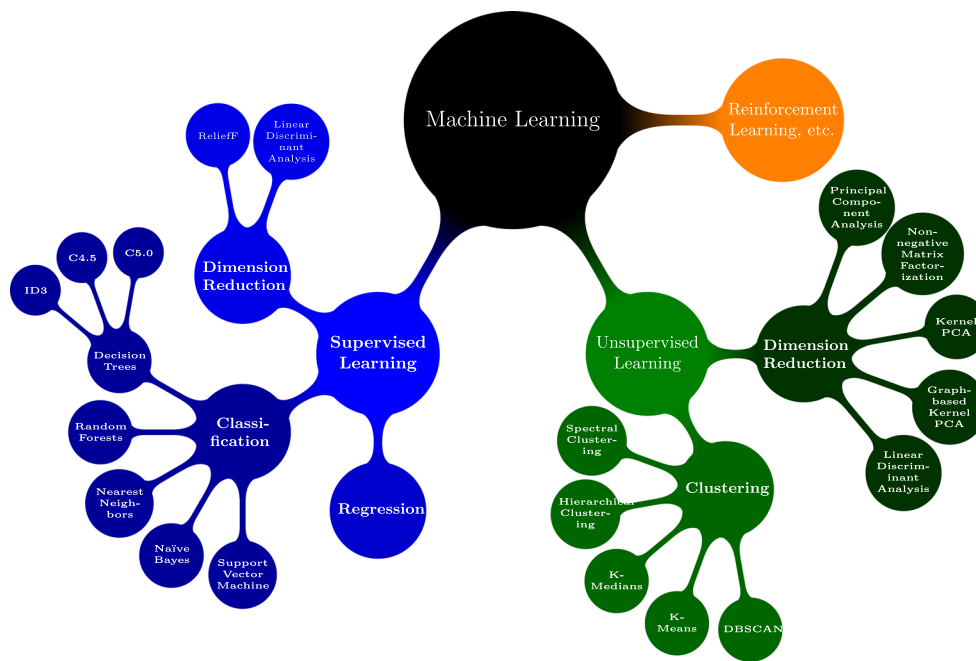
**Fig. 1** A Schematic Illustration of the Taxonomy and Example Algorithms in ML in Accordance with the Works of Russell and Norvig (2016) and Bishop (2006)

"Results" presents the results of our bibliographic review and subsequent scientometric analyses. Subsequently, we put these findings into context in "Discussion" and derive key challenges to wide-scale adoption and application of ML as a research instrument in IS. On this basis, we proceed by presenting the implications of our findings for IS research and practice in "Implications for research and practice". In the last step of our analysis, we derive a research agenda for applied ML research in IS in "Further research". The concluding "Conclusion" outlines the limitations of our research and offers additional suggestions for further research and a summary of our findings.

## Theoretical background

There are various definitions[1] for the research field and the term AI. In general, it can be described as the science that deals with intelligent agents that have set goals and interact with their environment by using their sensors and actuators to achieve these goals (Russell and Norvig 2016). By using their sensors, these agents perceive the environment to update their internal state, similar to how humans use their

organs to perceive their environment. In turn, these artificial agents use their actuators to assert an influence on their environment (Russell and Norvig 2016). Over the past 70 years, the research discipline of AI has developed different methods and directions of research that deal with AI in functional fields. Because the environment in which humans act contains a plethora of various influencing factors, data with different and changing characteristics, research has narrowed down the environment of AI algorithms to have a specific task within a particular environment. Therefore, in current research, AI is sometimes classified into two categories, "strong AI" and "weak AI" (Wayne and Pasternack 2011, p. 1224). "Strong AI" stands for a form of machine intelligence that today would be considered a utopia, as it would have to be equal to human intelligence. It can be inferred that all existent approaches involving AI can be classified as "weak AI". However, this distinction is not entirely uncontroversial. From our point of view, strong AI only represents a theoretical reference point. This makes it difficult to draw a clearly defined line. An alternative is a distinction between general AI and functional AI. As the current research suggests, ML excels with its statistical approach to problem-solving in these narrow, functional fields. The field of ML deals with the development of and research on computer-based algorithms capable of learning to improve their task performance (Jordan and Mitchell 2015, p. 255; Mjolsness and DeCoste 2001, p. 2051). A variety of ML algorithms have been developed to cover diverse problem types that were revealed

---

[1]For the sake of completeness, we must point out that by now, many different research streams have created an own understanding of the term AI. Therefore, various definitions have emerged that are suitable for a particular research area. However, we focus on more basic definitions that are closest to the original discipline from which the term originates.

**Table 1** Description of the most popular algorithms in the (German) IS community as assessed in the nationwide IS scholar survey in this paper

| | | | |
|---|---|---|---|
| Logistic Regression | Classification | Verhulst (1883) | Logistic regression is used to determine the relationship between one or more independent variables and a dependent binary variable. |
| Decision Tree | Classification | Belson (1959) | Decision trees are directed and ordered paths to represent decision rules (Kamiński et al. 2018, p. 138). |
| k-Means | Clustering | Lloyd (1982) | In this technique, a predefined number of groups is formed from a set of similar objects. |
| Nearest Neighbors | Classification | Altman (1992) | This technique determining the point in a given collection of data points that is nearest to a given query point (Beyer et al. 1999, p. 218). |
| Principal Component Analysis | Dimensionality Reduction | Pearson (1901) | This procedure transforms data sets by reducing the number of variables which static transformation so that in the end, a selection of significant linear combinations (main components) remains. |

through the formalization of real-world problems into computational problems (Jordan and Mitchell 2015, p. 255; Shafique and Qaiser 2014, p. 218). This circumstance leads to the problem that even the selection of appropriate models requires an intimate understanding of the data and knowledge of available ML tools, their properties, and the computational problems that they can (approximately) solve (Reicha and Barai 1999, p. 257). Therefore, an applier of ML methods first requires an overview of the central models of ML.

In general, ML methods can be segregated into two high-tier categories:[2] Supervised Learning and Unsupervised Learning. Figure 1 shows a taxonomy of machine-learning methods and corresponding examples of algorithms in Table 1).

Supervised learning uses features (i.e., attributes of the data of concern) to predict or classify a correct label (i.e., target value or ground-truth) (e.g., Hutson 2017, p. 19). Supervised learning models are generally trained on labeled data and make statistical inferences based on a trained model. In this regard, supervised learning models can serve the task of classification or the task of prediction. Well-known algorithms for classification are for example: Hierarchical Naïve Bayes (Wang et al. 2011), Decision Tree algorithms (van Riessen et al. 2016), Random Forest

(Sokolova et al. 2017), Support Vector Machine (SVM) (Piri et al. 2018).

In unsupervised learning, the algorithm looks for patterns in the features of a set of data while not taking ground-truth labels into account, if they are present at all (Hutson 2017, p. 19). In this regard, unsupervised learning allows for knowledge discovery even from unlabelled data. This pattern-detection may have various goals, for which different subfields of unsupervised learning exist. Clustering constitutes one of these subfields and can be defined as a procedure that groups data points within large data sets by similar properties (Gerber and Horenko 2015, p. 1). The models of these algorithms are not trained on previously classified data, such that the algorithms are bound to find patterns within the processed data. Well-known algorithms in the context of clustering are for example K-Means (Shahrivari and Jalili 2016), DBScan (Junior and da Silva 2017), Self-Organizing Maps (SOM) (Valle et al. 2017), BIRCH (Lorbeer et al. 2018) and Topic model (Gong et al. 2018). Another subfield that is often regarded as part of unsupervised learning is dimensionality reduction. The goal of dimensionality reduction is to discover compact representations of high-dimensional data by reducing the number of focal variables (Roweis and Saul 2000, p. 2323).

An essential sub-category of methods applicable in each subfield within ML is Artificial Neural Networks (ANN). Especially in the past decade, major improvements with respect to accuracy in automated problem-solving for complex computational problems such as the classification within the MNIST dataset (Deng 2012), would not have been possible without the ever-increasing availability of data and computational power. Because ANN may be the

---

[2]We opted to present some higher-order classes of algorithms as a single item to gain a good overview of important ML algorithms and to ensure the attentiveness of the survey participants. Therefore, our focus lies on Supervised and Unsupervised Learning. There are many different important ML subfields (e.g., Reinforcement Learning) that scholars of different backgrounds may or may not consider to be on the same level as Supervised and Unsupervised Learning.

recent powerhouse of efficient solutions to many real-world problems, the current hype around ML may have contributed to a large extent to the breakthrough of ANN, especially of Deep Neural Network (DNN)[3] architectures.

## Methodology

This section describes our research process in detail. This paper's core is a structured identification, evaluation, and analysis of relevant publications to answer the specified research questions. Our intention is not to provide a historical review that summarizes all of the ML papers in IS. Instead, we aim at answering research questions RQ 1.1 and RQ 1.2 by analyzing the emergence of literature involving ML methods in IS and the developments within IS literature over the past ten years ("Bibliographic review") and by merging the results of this literature analysis with findings from a nationwide survey of IS scholars ("Survey analysis").

### Bibliographic review and scientometric analyses

To answer our research questions, we make use of scientometric analyses based on a bibliographic review (Leydesdorff 2001). Scientometrics can be understood as the (quantitative) study of research, which provides information about publication patterns and progress of a field (Lowry et al. 2004). Especially in the light of the assessment of the current state of ML research in IS and further development of the discipline, scientometric analyses can yield systematic and analytic insights that may help to critically reflect the field and move it forward. Several senior IS scholars have deemed scientometric analyses as appropriate and necessary evaluation measures which can help confirm assumptions about inefficiencies in a research discipline and ultimately contribute to improving a discipline's standing (Lowry et al. 2004, 2013; Lewis et al. 2007; Straub 2006).

We, therefore, aim to conduct several scientometric analyses, encompassing publication patterns in IS journals, cooperation patterns on ML publications in the IS field as well as the utilization of ML methods. To perform insightful analyses, we must therefore employ a comprehensive bibliographic review that will provide the basis for our scientometric analyses.

A bibliographic review should follow established guidelines, especially if it deals with novel methods (Webster and Watson 2002). Therefore, we followed the guidelines for bibliographic reviews in IS proposed by Webster and Watson (2002) and Kitchenham (2004), and Wolfswinkel et al. (2013). As the first step, the guidelines propose choosing suitable selection criteria for the literature search. In our context, this includes the selection of keywords that either represent important and well-known machine learning algorithms or methods. Table 2 shows the relevant keywords. Subsequently, the next task is the selection of possible databases. We limited ourselves to a well-known selection of databases. These include: "AIS electronic library (AISeL)", "Google Scholar", "INFORMS database", "ScienceDirect", "SpringerLink" and "Taylor & Francis Online". We will describe these in detail later in this work. Step three is the actual search for literature, including querying the selection of databases with the selected keywords and search parameters. The final step comprises the review of the literature. This is the initial exclusion of non-relevant literature and the selection of relevant literature based on the title, keywords, abstracts, and articles' content. Based on our research question, we have selected the most important journals in Information Systems based on the definitions of the Association of Information Systems (AIS). At the current state, the AIS constitutes "the premier professional association for individuals and organizations who

**Table 2** Review process

| Publication from | 01/2010 - 12/2019 |
|---|---|
| Keywords: | Artificial Neural Network, BIRCH, Bayes, Classification, Clustering, DBScan, Decision Tree, Decision with Machine Learning, Deep Learning, Machine Learning, Naïve Bayes, K-Nearest Neighbor, K-Mean, Predictive Analytics, Random Forest, Regress, Reinforcement Learning, Supervised Learning, Self-Organizing Map, Support Vector Machine, Topic Model |
| Journals: | - European Journal of Information Systems (EJIS) |
| | - Information Systems Journal (ISJ) |
| | - Information Systems Research (ISR) |
| | - Journal of Association for Information Systems (JAIS) |
| | - Journal of Information Technology (JIT) |
| | - Journal of Management Information Systems (JMIS) |
| | - Journal of Strategic Information Systems (JSIS) |
| | - MIS Quarterly (MISQ) |
| | - Electronic Markets (ELMA) |
| | - Decision Support Systems (DSS) |
| Search engines: | Google scholar, INFORMS database, ScienceDirect, SpringerLink, tandfonline and journal database |

---

[3]The distinction between ANN and DNN is not trivial. A further examination of this topic can be found, for example, in the work of Chollet (2018) or in LeCun et al. (2015).

lead the research, teaching, practice, and study of information systems worldwide.".[4] In its role as a leading organ of the IS discipline, the AIS has also defined a list of top journals of the IS discipline, which is limited to those in the IS field.[5] Similar to other works that analyze the IS field (e.g., Trieu 2017; Khan and Triers 2018), we use the Senior Scholars' Basket of Journals as a basis to assess the current state of the IS discipline through analysis of its core journals. We add Electronic Markets (ELMA) and Decision Support Systems (DSS) as proxies for the most important journals in our research domain related to to Computer Science topics. Table 2 shows this review focuses on papers in any of the top or high-ranked IS journals (the Senior Scholars' Basket, plus ELMA and DSS).

The selection of the journals correlates with rankings such as the VHB-Ranking (German Academic Association for Business Research) or the CABS-Ranking (Chartered Association of Business Schools), which may be understood as a quality assurance measure for scholarly publications (Mingers and Yang 2017, p. 323). Thus, the selection process of the articles implicitly assumes quality assurance within the selected journals. Notably, this assumption is not undisputed according to Lehmann and Wohlrabe (2017) or Seiler and Wohlrabe (2014).

At this point, it should be mentioned that this quality assurance can be seen critically in the context of more technical research in IS, which includes the application of more complex ML methods. IS scholars who are rejected within the review process in one of these outlets that are considered as high-quality outlets by senior IS scholars (i.e., they are not accepted within the prevailing IS paradigm) proceed and often get the papers published in Computer Science journals, such as IEEE Transactions or in journals linked to the IEEE Computational Intelligence Society; these journals have much higher impact factors and are from a scientometric perspective (Leydesdorff 2001; Khan and Wood 2016) as relevant as the Basket journals but may (falsely, e.g., Lowry et al. 2004, p. 36) not be judged as strong contributions to IS research. Regarding data consistency and relevance across the sample, only publications containing the keywords from Table 2 in their abstract, title, or main text, were retrieved and analyzed. To validate that these rules were met, we used a two-layered approach, out of which an algorithm performs the first one, and a manual inspection is used as the second layer. To automatically query the databases, we developed a proprietary application.[6] After the .pdf files had been

retrieved by querying the databases and downloading the files corresponding to the articles, we used our algorithm to analyze these .pdf files. . Our algorithm opens all Portable Document Format (.pdf) files for the first validation step and tags the files that contain our selected keywords. In the second validation step, we manually inspected the tagged papers to check the classification's validity.

## Survey among IS scholars

As an additional means of analysis, a semi-quantitative survey is used to understand better how ML as a research instrument is perceived among the IS community. All IS researchers from German IS faculties were invited to participate in this survey. The questionnaire asks general questions on career development and specific questions on the subject area of the ML, especially the perceived transparency (or opaqueness) of a range of ML methods. Relevant findings are described in the second part of the results. The complete questionnaire, as well as the detailed visualizations of the statistics, can be found in the Appendix.

## Results

### Bibliographic review

We conducted the bibliographic review between July 2018 and July 2020. The initial results of our database queries included 1.922 papers. After file filtering by keyword matching using a proprietary algorithm, we retained 493 remaining papers. We then manually inspected the full text of the remaining publications. As a final result of the bibliographic bibliographic review, we ended up with a sample of 441 remaining papers. The bibliographic review of journal articles shows only a small number of contributions pertaining to appropriate keywords (see Table 2 for the list of keywords). Table 3 shows the results of literature research and scientometric validation, broken down by years and the journals examined.

The literature analysis shows a sobering picture concerning the journals within the Senior Scholars Basket. Only a few articles in these journals deal directly with the application of ML methods. In the top-ranking IS journals, such as ISR and MISQ, only a handful of scholars apply ML in their research projects. Exemplary among these are the works of Gong et al. (2018), Arazy et al. (2016) and Meyer et al. (2018). In contrast, the ELMA journal counts over twice as many publications as ISR. ISJ appears to be the only high-ranking journal to publish a substantial number of research articles that deal with the application of ML methods over the last ten years. At this point, it should be mentioned

---

[4]Cited from: https://aisnet.org/page/AboutAIS.

[5]Cited from: https://aisnet.org/page/SeniorScholarBasket.

[6]For more information on the algorithm, please refer to Appendix A.1, where the search algorithm is explained.

**Table 3** Reviewed journals and number of articles by year

| Journals | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | Sum | Freq. | Norm. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reviewed Journals | | | | | | | | | | | | | | |
| EJIS | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 3 | 6 | 0.5 |
| ISJ | 4 | 3 | 6 | 5 | 9 | 5 | 13 | 14 | 13 | 6 | 12 | 90 | 6 | 15.0 |
| ISR | 0 | 2 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 2 | 0 | 9 | 4 | 2.3 |
| JAIS | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 7 | 12 | 0.6 |
| JIT | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 1 | 0 | 7 | 4 | 1.8 |
| JMIS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 2 | 7 | 4 | 1.8 |
| JSIS | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 10 | 4 | 2.5 |
| MISQ | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 4 | 2 | 2 | 0 | 12 | 4 | 3.0 |
| EM | 1 | 1 | 0 | 2 | 2 | 1 | 1 | 2 | 5 | 5 | 4 | 24 | 4 | 6.0 |
| DSS | 15 | 15 | 19 | 22 | 20 | 26 | 28 | 27 | 29 | 38 | 33 | 272 | 12 | 22.7 |
| Sum | 20 | 24 | 28 | 30 | 34 | 36 | 44 | 52 | 57 | 61 | 55 | 441 | | |

Freq.: Is the publication frequency of each journal in each year

Norm.: Is the normalized number = Sum divided through Freq

that the journals are published at different intervals. Therefore, we normalize our results according to the publication frequency of each journal. Table 3 shows the normalized number in the "Norm." column.

Against this background, the number is to be regarded as relative. However, we argue with the absolute number to allow comparability with other literature analyses in IS. Notably, yet not entirely surprising, is the number of publications within the DSS journal with an application of ML methods: We found that 272 articles have been published over the past ten years within this journal, which is three times as many publications as in the ISJ. Notwithstanding, Table 3 shows a trend over the last years in the publication rate of ML in high-tier IS journals. From 2009 until 2019 the publication rate has increased approximately by 4.2 ($\beta$ : 4.2182 at $p < 0.01$) papers each year based on OLS regression analysis in Table 3.[7] This result indicates a growing interest in ML within the IS research domain. On closer inspection, journals with a more technical focus like DSS, or top-journals like ISJ, that state "publishing high quality, yet risky, papers that help to move the IS field forward" (Davison et al. 2012; Davison 2017) as part of their mission and tradition are highly benefiting from this growing interest.

Nonetheless, it seems that it has been difficult in the past to publish research articles that apply ML methods as research instruments in top IS journals according to our results. Similar observations concerning the application of

ML in research have been made by Krauss et al. (2017) in the neighboring discipline of finance. The stagnation in applied ML research may be a problem that is originated in the reliance of IS research on standard quantitative methods, which is strongly influenced by organizational studies (Basden 2010). Based on the assumption that the disciplines influence each other, we argue that additional information about the collaborations and affiliations of researchers to different research areas within our sample would facilitate understanding. In line with Khan and Wood (2016) we argue that collaborative behavior influences the research diversity with all its virtues and thus influences the content, performance, and output of those involved in its boundaries (Vidgen et al. 2007).

Therefore, we analyze the collaborating research disciplines in applied ML research in IS in the next step. To classify our results, we analyzed the collaborating disciplines within our sample of 441 research articles. For this purpose, we extracted the additional information regarding the authors of the articles from their websites, classified their affiliation, and organized the information on collaboration within a matrix. Figure 4 shows a clear pattern regarding the collaboration structure for papers that apply ML in IS research. Of the 441 papers, 224 have been published by heterogeneous teams. Homogeneous teams have published 217 papers (in relative terms: 38% Information Systems, 25% Management Science, 13% Computer Science, 11% Engineering, 5% Industry research, 8% remaining research fields, see Fig. 4).
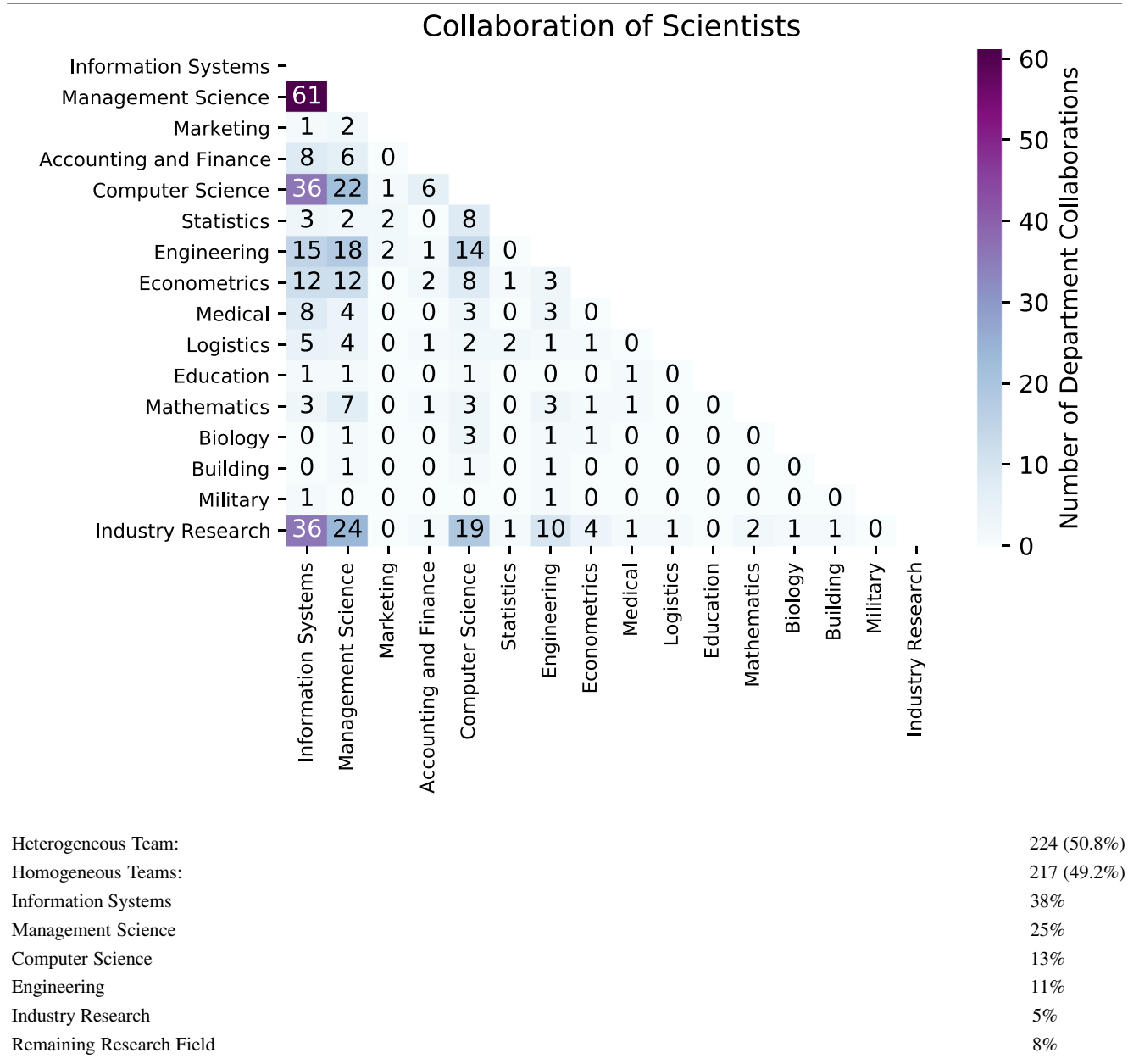
To identify the collaborations, we have counted each connection in the teams once, which means that if several scholars of the same discipline research area are listed in the article, it was counted as one occurrence. We discovered

---

[7] Ordinary Least Squares (OLS) regression with time and publications number: $y = \alpha + \beta * x_t + \varepsilon_t$; $\alpha$ : 18.8182***, $\beta$ : 4.2182***, Adj. $R^2$: 0.938, *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$, $\alpha$: Regression intercept, $\beta$: Growth rate of ML related publications, $x_t$: Publication year of ML related publications.

strong scientific collaborations between IS scholars and scholars from Management Science, Industry research, and Computer Science. The finding shows a congruency with the ideas of Basden (2010) that IS research tends to have more collaborations with departments from business schools and management science, rather than departments from its other neighboring discipline, Computer Science. Noticeable in our analysis in the context of ML is the high participation of companies. Altogether, IS journals represent a large number of disciplines in the context of ML. We also discovered exotic papers such as Swiderski et al. (2012) from the field of military research.

To give a clear and informed answer to the question of which analytical research instruments dominate IS research, we decided to employ a trend analysis and compare other statistical methods used in IS research. For this reason we developed a data mining algorithm to extract all necessary information from the "Association for Information Systems" (AIS) online library. In line with Recker (2013) and Chen and Hirschheim (2014) IS research can be systematized into three categories: quantitative, qualitative and mixed (Recker 2013, p. 66). These categories are an appropriate starting point because IS scholars use diverse methods and theoretical lenses to

**Table 4** Scholar affiliation/collaboration matrix in is research works with ML methods



| | Information Systems | Management Science | Marketing | Accounting and Finance | Computer Science | Statistics | Engineering | Econometrics | Medical | Logistics | Education | Mathematics | Biology | Building | Military | Industry Research |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Management Science | 61 | | | | | | | | | | | | | | | |
| Marketing | 1 | 2 | | | | | | | | | | | | | | |
| Accounting and Finance | 8 | 6 | 0 | | | | | | | | | | | | | |
| Computer Science | 36 | 22 | 1 | 6 | | | | | | | | | | | | |
| Statistics | 3 | 2 | 2 | 0 | 8 | | | | | | | | | | | |
| Engineering | 15 | 18 | 2 | 1 | 14 | 0 | | | | | | | | | | |
| Econometrics | 12 | 12 | 0 | 2 | 8 | 1 | 3 | | | | | | | | | |
| Medical | 8 | 4 | 0 | 0 | 3 | 0 | 3 | 0 | | | | | | | | |
| Logistics | 5 | 4 | 0 | 1 | 2 | 2 | 1 | 1 | 0 | | | | | | | |
| Education | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | | | | | | |
| Mathematics | 3 | 7 | 0 | 1 | 3 | 0 | 3 | 1 | 1 | 0 | 0 | | | | | |
| Biology | 0 | 1 | 0 | 0 | 3 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | | | | |
| Building | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| Military | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| Industry Research | 36 | 24 | 0 | 1 | 19 | 1 | 10 | 4 | 1 | 1 | 0 | 2 | 1 | 1 | 0 | |

| | |
|---|---|
| Heterogeneous Team: | 224 (50.8%) |
| Homogeneous Teams: | 217 (49.2%) |
| Information Systems | 38% |
| Management Science | 25% |
| Computer Science | 13% |
| Engineering | 11% |
| Industry Research | 5% |
| Remaining Research Field | 8% |

explore the phenomena of research interest. We add a fourth category, namely ML. This enables us to analyze the papers according to the classical methods, such as regression (Goodhue et al. 2017, p. 667) or partial least squares regression (Marcoulides et al. 2009, p. 172). Within these four categories, our algorithm assigns the applied research method to its corresponding category. To assign the right category for each used research method, we use the work of Chu and Ke (2017) as guidance. In line with Recker (2013) and Chu and Ke (2017), this analysis shows a similar relative trend for quantitative, qualitative, and mixed methods. The depicted analysis is both congruent with our findings in Tables 3 and 4 and with the idea of Basden (2010), that IS is influenced by a variety of different disciplines. While Fig. 2 as well as Table 3 show a rise in the use of ML methods, these publications seem to be mostly accepted in conference proceedings or 2nd-tier journals, such as DSS and EM, but not in journals from the senior scholars' basket of journals, such as JIS, ISR or MISQ. Although the philosophical notion of "spheres of meaning" take on our discipline by Basden (2010) is certainly disputable, it may explain why cutting-edge IS research falls behind in the application of more recent and probably more advanced analytical and data-driven methods.

## Survey analysis

With regard to Basden (2010), it appears that the major concern of disciplines from organizational studies is not the development of stronger different analytical or data-driven approaches, but the assurance of the validity of social, economic and formal theories (Basden 2010). Since IS research has its roots in organizational studies, the quantitative approaches that are applied are the ones that have been accepted as valid for a long time, which may make it more difficult for newer methods. One possible cause of the lack of application of certain ML methods is their black box characteristic, which means that the processes between input and output are opaque, such that it obstructs intuitive interpretability of a method. As put forth by a senior IS scholar lately, "inscrutability can hamper users' trust in the system, especially in contexts where the consequences are significant, and lead to the rejection of the systems."(Rai 2020, p.1). To verify our notion about the connection between a lack of interpretability and a lack of applied ML research in IS, and to assess the perceived black- or white box-characteristics (i.e., intuitive interpretability) of the applied ML methods by IS scholars, we conducted a short semi-quantitative survey with IS research scholars in Germany. The invitation was sent out to all members of German IS faculties. At this point, it must be emphasized that education systems in different countries in the IS community deal differently with ML, which is why the sample can be considered indicative only. We asked 364 scholars and received 110 answers (response rate of 30%). We let IS scholars rank different ML methods on a scale of 1 to 5 (1 = White box to 5 = Black box).

Regarding the demographics, most participants in the survey are researchers from the IS field, male and working as research assistants. With respect to their research orientation in the context of the area of highest education, the largest group of participants have studied IS with 68.25% (accounting and finance 1.59 %, econometrics 1.59

**Fig. 2** Trend of Research Methods Used in Publication in Selected Top IS Conferences and Journals Vs. Machine Learning as Instrument. Time frame of the publications 01/2009 - 12/2019
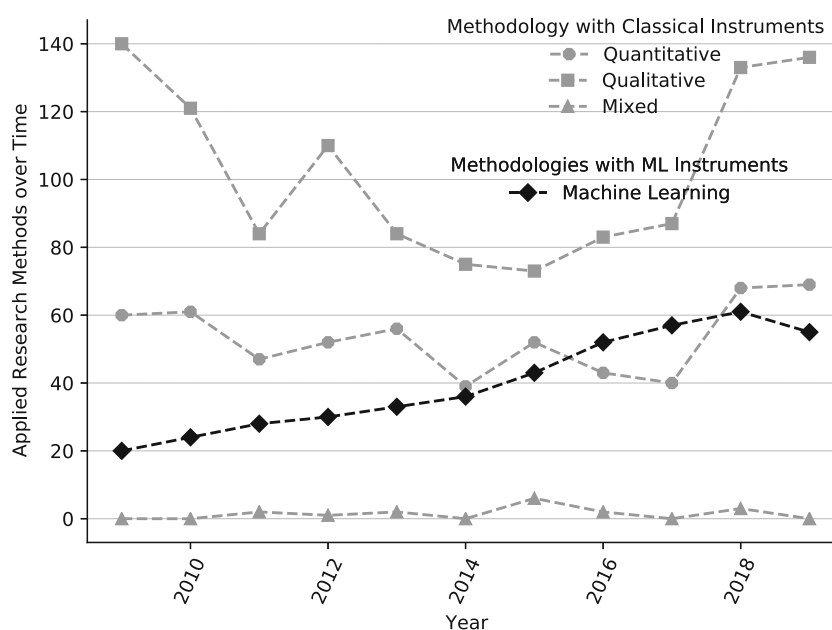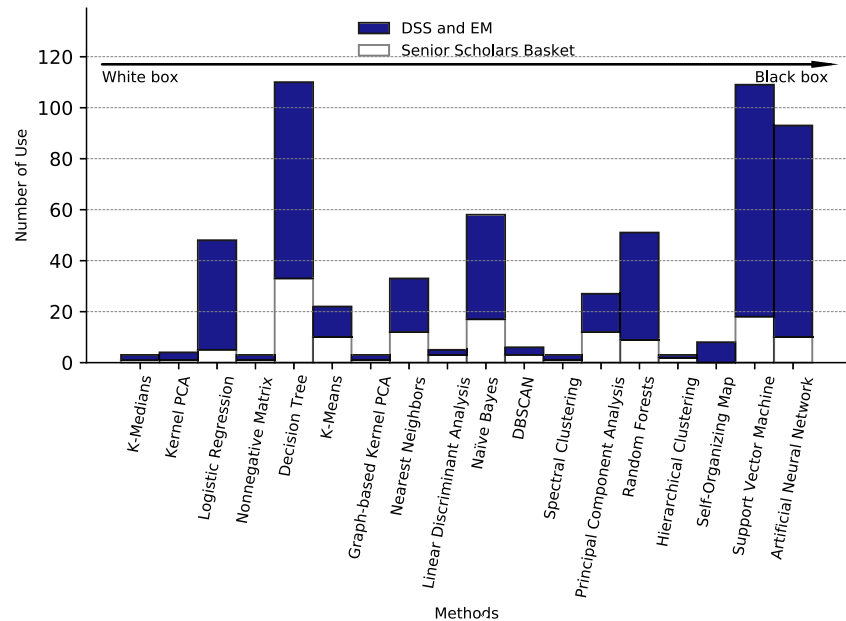
**Fig. 3** Usage of ML Methods in IS Publications. The methods are ordered according to their degree of the black box characteristics (top = White box method, bottom = Black box method)



%, logistics 3.17%, statistics 3.17%, engineering 4.74%, other 4.74%, management (science) 4.74%, Computer Science 7.93%). The largest proportion is male with 85.7% (female 12.7% and other 1.59 %). The survey was mainly filled out by Doctoral or PhD candidates 65.1% (Prof. Dr. 1.11%, research assistant (without being Dr. or PhD candidate) 20.65% and Other 3.17%).

The data collected reveals that 93.2% of the participants did not have dedicated ML training courses in their PhD program. However, 74.1% say that they would like to have a training course in ML.

In line with our RQ 1.1, we asked about the relevance of ML in IS. In response, 85.4% of researchers stated that ML is absolutely relevant. Furthermore, 77.4% of the researchers also assume that ML is relevant or absolutely relevant for the progress of the IS discipline. Also, 85.4% of researchers considered the relevance of ML for IS research to be persistent in the future.

We continued our analysis by calculating a score measuring the degree of the black box characteristics based on the IS researchers' assessment. Figure 3[8] shows ML methods sorted according to their transparency (K-Medians = White box; ANN = Black box) based on the IS researchers' assessment.

Figure 3 shows that it is generally accepted that ANN are considered a black box. On the other hand, Decision Trees are more likely to be rated as a white box method.

The comments from the open text fields of the survey reveal that Decision Trees are seen as a white box because they provide metrics that are simple to interpret (e.g., entropy, information gain). Our investigation shows that

most of the works in our original sample (N = 441) that are published in the Senior Scholars Basket of Journals deal with simple and well-known approaches like Naïve Bayes and Decision Tree algorithms. Approaches with lower transparency, such as ANN, BIRCH model, or SOM, are used less often or not at all.

To determine the publication rate in top journals in relation to the method complexity level (1 = White box to 5 = Black box), we performed a logistic regression. In the case of multiple applied ML-methods, the method with the highest degree of non-transparency is considered in our analysis.

$$\text{Publication in Top-Journal}(y_{1|0}) = \alpha + \beta_1 \text{Complexity} + \varepsilon \tag{1}$$

The odds ratio (0.8193) indicates that if the degree of the opaqueness of the applied ML method increases by one unit, the probability of a publication in a top journal decreases by 18.07%. These findings are in line with many overview papers on ML in the field of IS, such as Chen et al. (2012), Dhar (2013), Abbasi et al. (2016), and Maass et al. (2018). Compared to more traditional statistical methods used in IS, such as partial least squares regression (PLS regression) or structural equation modeling (SEM), ML methods are relatively sophisticated (Table 5).

In fact, advanced ML models can capture any functional relationship between variables given enough data (Pearl and Mackenzie 2018). Although the capabilities of ML methods often go beyond what is achievable with traditional methods, a lack of transparency leads to problems with the intuitive interpretability of ML methods. Due to this circumstance, it is not a surprise that ML methods are

---

[8]The detailed analysis can be found in the appendix in Table 8.

**Table 5** Logistic regression for acceptance determination in top-journals considering the ML-Method complexity

| Logistic regression | | | | | |
| --- | --- | --- | --- | --- | --- |
| | Independent variable | Std. Err. | t-Stat. | *p*-Value | Odds-Ratio |
| Constants | − 0.3755 | 0.256 | − 1.465 | 0.143 | 0.6869 |
| Complexity | − 0.1992 | 0.082 | − 2.418 | 0.016 | 0.8193 |

Pseudo R:0.01680, Log-Likelihood: − 137.92, Likelihood Ratio p-value: 0.01478

not widely used in IS research, which supports our hypothesis that the black-box characteristic and a lack of interpretability may be reasons for the low application rate of complex ML methods in top IS outlets.

## Discussion

Our analysis revealed only a small number of articles that apply ML methods in IS top journals. The number of articles dealing with ML in the IS domain has increased over the years, which suggests that this is a research area that is increasingly attracting academics' interest, especially in the past two years. We share the concern "that the IS research community is making the discipline's central identity ambiguous by, all too frequently, under-investigating phenomena intimately associated with IT-based systems and over-investigating phenomena distantly associated with IT-based systems" (Benbasat and Zmud 2003, p. 183).

One reason for the findings of RQ 1.1 may be that articles that are concerned with or built upon ML methods' application are rather unpopular among scholars and were excluded in the review process. Due to the low prevalence of ML methods in IS research and the associated low confidence in the robustness, verifiability, and interpretability of the results, a systematic negative bias towards these methods could exist. Due to the perceived lack of insight, articles dealing with or involving ML topics could be difficult to publish in high-ranking journals because articles with traditional methods are preferred. Thus, authors may only be motivated to use classical quantitative methods, as they provide an easily justifiable method and allow a direct result in terms of knowledge gained with a transparent white-box approach.

Another possible reason for the findings related to RQ 1.1 may be, of course, that a systematic lack of expertise and understanding of ML methods is present in the domain of IS. This circumstance may be directly related to the analysis in the previous paragraph: If acceptance rates, reviewers or PhD programs tell IS scholars that other research instruments are not equally valuable, researchers will not strive to learn and apply these new methods

strenuously and will prefer to use classical quantitative research methods. In addition, our survey findings support the idea that only a few PhD students receive appropriate education in ML methods. From our participants, only 6.78% of the PhD students completed required ML courses during their PhD (see appendix). Consequently, prospective researchers in our field have hardly any contact with these technologies during their training.

In contrast, our survey data indicates a clear interest in such courses (74.07% of the participants would like to attend such courses). The integration of ML courses into regular IS-PhD programs could solve this problem and help researchers learn to apply these methods correctly in their early careers. Nevertheless, based on our quantitative data, we cannot tell if researchers, in general, perceive ML methods as hard to learn, hard to apply, not as valuable, or not as interesting as standard quantitative methods.

## Analyzing the central challenges from the KDD perspective

We draw the notion from our scientometric literature analysis results that the underlying issue warrants further qualitative examination. Using the acquired information from our analyses, we identified essential obstacles and challenges for large-scale adoption of ML methods within IS research. We structured insights from the reviewed literature into a total of three central challenges that IS scholars, as well as industry practitioners, should be aware of if they aim to leverage the benefits of ML in research or practice. In principle, these challenges for the application of ML in the field of IS can be systematized along the Knowledge Discovery in Databases (KDD) model (e.g., Fayyad et al. 1996).

In general, KDD can be understood as a process guideline for information analysis and knowledge extraction, which is divided into stages from data acquisition to data interpretation. The objective of KDD is knowledge inference such that an end-user may finally gain valuable information that supports them in their attainment of a guideline for the development of methods for making sense of data (Fayyad et al. 1996, p. 37) is also a part of the Data Mining (DM)

process (Marban et al. 2009). The KDD model expects the use of input data derived from a dataset acquisition process (Lara et al. 2014, p. 54). The raw data are filtered according to quality criteria, selected according to possible target values, and transformed correspondingly. These essential steps can also be referred to as pre-processing, feature selection, and feature engineering. Subsequently, the selected Machine Learning model will be trained with the prepared data. The interpretation of the data by a user follows and a recursive-iterative process for the fine-tuning of the model may occur until the final model is determined.

Although the KDD model poses a helpful guideline, many challenges emerge along with the respective process steps, which we will now address according to the three central challenges we identified: i) data preparation, feature selection, and engineering, ii) model optimization and parameter tuning, iii) results and interpretation of black boxes.

## Data preparation, feature selection, and engineering

ML methods heavily rely on high-quality training data. The work of Baier et al. (2019) shows that for many applications of ML models in practice and science, one of the biggest challenges is data quality (Lee-Post and Pakath 2019). Various factors may determine if the quality of the data to be used is sufficient for the project in pursuit. Among these factors, the properties, structure, and complexity of data samples (Piri et al. 2018, p. 23) may play a huge role while also contextual factors, for example, timestamp of data collection could have an impact on the results. Data generally tend to exhibit noisy parts. That is why data screening, cleansing, and pre-processing need to be performed before ML algorithms can use these data as an input (Basti et al. 2015, p. 22). One major problem resulting from data cleansing is the possibility of losing important information by removing certain parts of the data (Rittgen 2009). As a result, variables generated from data mining models could lead to biases, and misclassifications (Yang et al. 2018, p. 4). Therefore, it is important to realize that it is common practice to dedicate about 80% of the labor and time within the KDD process to data preparation, as a representative survey on a popular data science portal shows.[9] This circumstance shows that there is potential for further research in the area of data preparation to reduce the time consumption of data preparation. This circumstance is linked to the quality of the data. Future research, such as in the field of Resource Description Framework (RDF) can contribute to this (e.g., Benbernou and Ouziri 2017).

Besides questions of data quality, privacy concerns play an important role. The relevance of this topic increased drastically from the early history of data privacy since the 1980s (Mason 1986; Bélanger and Crossler 2011). Since these beginnings, the relevance of privacy has increased significantly in theory and practice along with the use and development of more efficient yet data-hungry information processing technology (Pavlou 2011). Especially modern methods from the field of ML can use large amounts of data and even have to do so to achieve sufficiently accurate and generalizable results. But the demand for information privacy leads to new challenges, especially for companies (e.g., Casey et al. 2019), which is expedited by the uncovering of numerous violations of privacy in the past (Culnan and Williams 2011), and the resulting public outcry for more transparency and more restrictive policy regulations such as the. GDPR[10] Calls for better integration of privacy and privacy-compliant IS research from almost ten years ago (Bélanger and Crossler 2011) are thus certainly more relevant today than ever before, especially regarding the usage of ML.

In addition to general quality concerns, the sample size is as important for ML algorithms as for traditional quantitative analysis methods because the generalizability of ML algorithms can in most cases only be granted by training on large sets of sample data (Goodhue et al. 2012, p. 983). Since the available data must be used to train the algorithm and test and validate it, it appears to be an equally important task to divide the available sample data into training and testing sets.

The complexity further increases when dealing with textual analysis or other NLP tasks, which imbues the problem of unstructured textual sample data (Wang et al. 2015, p. 90). In the case of NLP tasks, the data has to be thoroughly cleaned during the pre-processing phase. For example, punctuation, numbers, and abstract structures such as hyperlinks should be removed before the actual analysis because hyperlinks and punctuation do not themselves add any new information, yet may create noise (See-To and Yang 2017). Also, the data must be continually analyzed to understand what implications the results of an analysis may hold in general (Cnudde and Martens 2015, p. 83). In conflict with this approach is the fact that a lot of real business data, which are provided and analyzed within research studies, can only be retrieved in an encrypted form so that humans can not form a deeper understanding of the data yet still feed their algorithms with it (Martens and Provost 2014, p. 884).

---

[9]https://bit.ly/2WwVPho.

[10]For more details, see Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

**Table 6** Transformation of the city feature to an ordinal numerical feature space, resulting in an implicit bias

| record | city | rating | | record | city | rating |
|---|---|---|---|---|---|---|
| 1 | Paris | High | $\rightarrow$ | 1 | 0 | 2 |
| 2 | Boston | Medium | | 2 | 1 | 1 |
| 3 | London | Low | | 3 | 2 | 0 |

While the availability or creation of high-quality data is already a good basis for an accurate predictive model, model performance may still be highly dependent upon the selection and engineering of appropriate features (e.g., Zhang et al. 2017). For example, in unsupervised and supervised learning, feature engineering (e.g., Au 2018) is important for efficient ML-modelling, since provided data - even if it is clean, may warrant unwanted biases[11] due to its mere structure (e.g., Feldman et al. 2015). Take, for example, two categorical features: city and rating. Both features contain a set of categorical values and both need to be transformed. Nevertheless, while transforming the "rating" feature into an ordinal vector space may lead to an appropriate result, the same procedure would result in an implicit bias for the "city" feature (see Tables 6 and 7[12]). Instead, the city feature should rather be one-hot-encoded to avoid such a bias. On the other end, procedures like one-hot-encoding result in a larger feature space, inducing the problem of the "curse-of-dimensionality", which again may lead to a detrimental degradation of model performance. This problem, on the other hand, should be tackled by dedicated feature selection procedures, that aim to select the most relevant features for the model and ML task at hand (e.g., Zhang et al. 2017; Bach 2017). While many feature selection approaches have been developed until now, research is still very active in this area, and novices may have a hard time finding an appropriate approach for their endeavor. Depending on the feature engineering and selection approaches the ML applicant vows for, model results may vastly vary, which increases the difficulty to determine if the selected methodology is appropriate or not, even for expert reviewers.

One additional aspect within the context of data preparation is the access to sources with relevant and interesting data, which may explain the high number of publications with industry collaborations (Fig. 4). Examples include collaborations with companies such as Google, Deloitte Consulting, and IBM (Lozano et al. 2017; Fu et al. 2017; Pai et al. 2014).

## Model optimization and parameter tuning

The use of ML algorithms frequently involves careful tuning of learning parameters and model hyperparameters (Snoek et al. 2012, p. 2951). Incoherence with the data, the preparation task is the challenge of model choice. Some scholars have examined this issue in their works, such as Gao et al. (2017, p. 36) or Evermann et al. (2017, p. 139). Whether to choose unsupervised or supervised learning methods depends on the problem environment as well as on the properties of the available sample data (Lau et al. 2012, p. 1245).

Only few papers use a combination or layering of several algorithms to test the robustness of their results (Martens et al. 2016, p. 75). This approach may be useful and appropriate for cases in which data needs to be processed in different phases to infer knowledge from the data. This implies higher complexity, which is a reason why such a multi-method approach is rarely used.

An important problem of model optimization is the risk of overfitting related to predictive models, which results in overoptimistic results (Siering et al. 2018). To reduce the problem of overfitting, the choice of a simple model can be a possible solution. Scholars and practitioners alike seem to think that the more complex an ML model is, the higher its predictive power and robustness may be. Nevertheless, this may be a fallacy, as pointed out by Cresci et al. (2015). Indeed, it seems that the more complex a model is, the more likely it is overfitting. This should lead us to rethinking this general paradigm of complex architectures or to perhaps turning to simpler, yet more effective models (Cresci et al. 2015).

With the high-tier categories of ML in mind, the choice of the ML learning method and algorithm depends on not only the problem environment but also the properties and structure of the available data. For example, available data could either be numeric or alphanumeric, discrete or continuous, structured or unstructured. Another important point is the choice of the right model parameterization - for example, choosing the right k for a model of the k-means algorithm (Li et al. 2017, p. 83) or the choice of the

---

[11]Whenever we refer to biases in data or models, we do not implicitly refer to notions of Algorithmic Bias in terms of disparate treatment, but rather to disparate impact or statistical bias. For more information on algorithmic bias, refer to the provided source.

[12]Depending on the model type and the task, it should be reflected if it makes sense to additionally drop one of the resulting columns of the one-hot-encoded feature.

**Table 7** Transformation of the city feature through one-hot encoding, avoiding implicit bias

| record | city | rating | | record | city_Paris | city_Boston | city_London | rating |
|---|---|---|---|---|---|---|---|---|
| 1 | Paris | High | $\rightarrow$ | 1 | 1 | 0 | 0 | 2 |
| 2 | Boston | Medium | | 2 | 0 | 1 | 0 | 1 |
| 3 | London | Low | | 3 | 0 | 0 | 1 | 0 |

number of layers, iterations, and batch-size of training data for an ANN. The parameterization may strongly influence the accuracy and explanatory power of the models (Walczak and Velanovich 2018, p. 117). In addition, the choice of training data is also a difficult task, for which procedures like k-fold cross-validation are trying to compute the suitable training set (Topuz et al. 2018, p. 102); Singh and Tucker (2017, p. 87) or the selection of the right parameters for Support Vector Machines from the available data sample (Huang et al. 2016, p. 22).

Model optimization and parameterization of ML models are, therefore, very challenging tasks. Beyond that, scientists and managers have to solve further problems if they want to apply ML methods. These challenges can be attributed to the lack of standardization, in the sense of theoretical guidelines for using ML methods on economic data. Possible questions that may arise include: How should the data be prepared and processed? Which methods are the most promising ones for the analysis of the underlying data? How should the parameters be set so that the model will achieve persistently verifiable and robust results?

We believe that a robustness analysis must answer these questions. In our review, we found that only a few papers conduct a robustness analysis of their models, such as Wang et al. (2015) or Scholz et al. (2016). Furthermore, few papers present the exact parameters used within their models like Pinto et al. (2015), which may be essential for result evaluation and verification by an independent reviewer or scholar. Such an analysis may become even more important when authors propose combining different methods, such as (Wan 2015). Therefore, scholars and practitioners alike should be encouraged to not only thoroughly analyze their problem when using ML methods but also to conduct robustness analyses and to disclose their approaches with respect to the parameterization of their models to increase transparency and trust in the reliability and robustness of their analysis.

## Results and interpretation: the black box

One of the main challenges for the large-scale adoption of ML in IS is the black box characteristics of most methods at run time (Müller et al. 2016, p. 294). Our results, (e.g., Fig. 8), support this notion. ML methods, especially ANN, are often understood and labeled as black box solutions because they provide little explanatory insight into the influence of the independent variables and on why the computations within the ANN lead to specific results (Olden and Jackson 2002, p. 135). Truly understanding why certain results are given may require applications of ML methods to dive deep into the training algorithms of the ML algorithms themselves. Especially with the advent of highly non-linear complex inference systems in the form of DNN, this problem has become apparent to scientists of many disciplines, to the extent that the task of unboxing the black box has even spawned dedicated workshops at leading ML conferences such as the ICML.[13] Several promising methods for interpretability have been developed until now, ranging from software packages for local, model agnostic explainability (e.g., Ribeiro et al. 2016) for a large range of models, over game-theoretic approaches to explainability (e.g., Lundberg and Lee 2017) to explainability methods, developed especially for ANN interpretability (e.g., Kokhlikyan et al. 2019). Since there are a lot of possibilities for including explainability methods today, we would strongly recommend using these methods. For one, they help the scientist create better models and improve the results and second, they may help increase the reviewer's trust in the methodology.

## Implications for research and practice

## Implications for IS research

Due to the outlined points, scholars face a huge challenge to publish ML papers in high-level IS-journals because the reviewers are aware of these problems. Although the submitted works could provide interesting results, they can be rejected by the reviewers because i) the expert sees a trade-off of rigor for relevance, ii) a deeper understanding of analytical procedures is difficult or impossible due to a black box characteristic, iii) traceability of the

---

[13]For example: 2018 Workshop on Human Interpretability in Machine Learning (WHI), https://sites.google.com/view/whi2018/home.

methodology is impossible due to data pre-processing, parameter tweaking or overfitting. We believe that this is one of the profound explanations for the small number of high-level publications containing ML methods in the IS research domain. On a different note, a publicly accessible scientometrics project that analyzed Computer Science publications between 2012 and 2017[14] revealed that 28,303 papers which dealt with topics in the field of ML were published in this timeframe alone. Computer Science deals with the improvement of algorithms, IS applies them (if they are contextually appropriate) to solve problems or gain new insights. We, therefore, hope that our paper provides a fresh impetus to researchers to not only increase their chances of acceptance by describing their methodology in using ML methods more thoroughly but also to improve our understanding of the results from ML in IS research in general, such that result reproducibility will lead to consequent cutting-edge research projects.

## Implications for practice

In principle, ML methods are promising for optimizing processes, products, services, or for supporting management decisions through inferential statistical analyses (Baesens et al. 2016, p. 810). However, as soon as the results of the ML methods have to be understood by third parties in the company who do not have the expertise in the statistical procedures on which the ML methods are based, the methods lose credibility and usefulness in practice due to the black box problem. This poses the central problem of the interpretability of the results for the application of ML methods in the business sector. For business managers, it is important to understand the recommendations; therefore, the results must either be self-explanatory or easily traceable (Bose and Mahapatra 2001, p. 215).

This circumstance is especially dangerous for middle-sized organizations and organizations with lower IT expertise. These organizations are under increasing pressure to develop technologies that allow them to compete within their market, especially with regard to the penetration of different industries by IT companies (Kohli and Melville 2018, p. 1). As a particular example that shows how traditional players can be threatened by IT-driven companies which focus on novel technologies and their technological expertise, the market disruption in payment transactions could be mentioned, which was disrupted by the ascent of

technology firms like Apple (Puschmann 2017, p. 69). To come back to the problems of the application of ML methods in the business context. On the one hand, the management of traditional industry companies do not embrace the technologies and do not completely tap their full potential.

On the other hand, high-tech companies like Amazon dedicate a lot of financial and human resources to the development of systems with ML applications, e.g., to increase satisfaction in customer service (Nwankpa and Datta 2017, p. 472). This discrepancy in the strategies of traditional industry players and IT-players may lead to an outflow of knowledge, the missing out on game-changing opportunities, and finally, the loss of competitiveness with all its economic consequences (Krauss et al. 2017, p. 48).

## Further research

The findings of our bibliographic bibliographic review suggest that IS research has been reluctant to apply ML methods within research articles, especially regarding publications in high to top-ranking journals. This also implies that a far-reaching fundamental understanding of the chances and risks of applying ML methods has not yet been developed within our research domain. Generally speaking, IS research needs to engage further with ML and its applications on socio-technical problems that we face in our research discipline, to understand the opportunities and risks in the innovation context. As a hyphenated research domain between management and Computer Science, IS research needs to face the challenges and opportunities of ML methods and provide this knowledge to other scholars in social sciences and other management disciplines.

Our results indicate that the majority of scholars in the IS research domain follow the classical (quantitative analysis) approach, which is founded on hypothesis-based research. Following that finding, an overarching challenge arises in IS because ML is often used as part of an outright different approach. Regarding the example of pattern recognition, a general idea may arise that certain sample data could provide new insights for a specific research case, yet a hypothesis may not be put forward unless the data has been analyzed and patterns have been found within the data. In consequence, and in line with Maass et al. (2018), matches between two different approaches that rectify the application of ML methods based on theory must be found.

Hence, an important challenge is that bridging the gap between the data-driven and the theory-driven research models requires researchers to engage in tasks for which no single individual may be perfectly equipped because they require knowledge and skills related to data analytics

---

[14]"A Peek at Trends in Machine Learning" by Andrej Karpathy, Source: https://medium.com/@karpathy/a-peek-at-trends-in-machine-learning-ab8a1085a106.

techniques, as well as expertise in relevant domain theories (Maass et al. 2018, p. 1259). Subsequently, further research needs to deal with the gap between hypothesis-based research and the pattern recognition approach for ML. The two approaches should be extensively tested and evaluated in business and industry applications where productive problem solving and decision-making with Computational Intelligence (e.g., Sha et al. 2019, p. 108), ML and Soft Computing methods (e.g., Ibrahim 2016, p. 34) can be set up, implemented, carried out and evaluated. Accepted, conservative (hypothesis-driven) methods may follow the IS paradigms but will be meaningless if they appear unable to handle real-world problems in a digital economy and Big Data context.

In consequence, IS research has to address several questions that allow a theoretical foundation for research to be built that embraces ML as a powerful research instrument in IS, and which coincides with the requirements of Maass et al. (2018) in the context of data-driven and theory-driven research:

1) Is it possible to amalgamate the hypothesis-based research and the data-driven ML approach?
2) How can we guarantee transparency in ML methods and models, such that the understanding, interpretability, and validity increases to achieve broad acceptance of these methods in the IS domain and business?
3) How can we use the benefits of the amalgamation of the two approaches in science and business?

4) How can we effectively integrate ML as a research instrument within qualitative, quantitative, and mixed-methods methodologies?

About our findings from RQ 1.1 and RQ 1.2, we compiled a preliminary Information Systems research framework based on the work of Hevner et al. (2004). Figure 4 shows our framework. This framework aims to tackle the previously mentioned four elementary questions to enable a broad acceptance and structured application of ML methods within IS research.

The research framework is based on the idea that IS research usually embraces business needs and develops theories from its existing theoretical knowledge base and aims to justify these theories by analyzing data sets using classical methods and instruments. In the justification process, again, the foundational knowledge and methodologies of ML may increase the richness and power of analytical instruments and provide groundbreaking insights, compared to classical analytical instruments. Furthermore, especially when trying to gather novel insights on phenomena that concern ML-based systems directly, it may be fruitful to modify such systems, instead of simply observing and analyzing the phenomena with ML-based systems as a distant and static IT-artifact. Having the possibility of modifying ML-based systems would potentially enable us to unravel different and more nuanced facets of this technology's behavioral influences.

Although one may argue that classical methodologies (i.e., case studies, Wizard of Oz experiments) may be power-
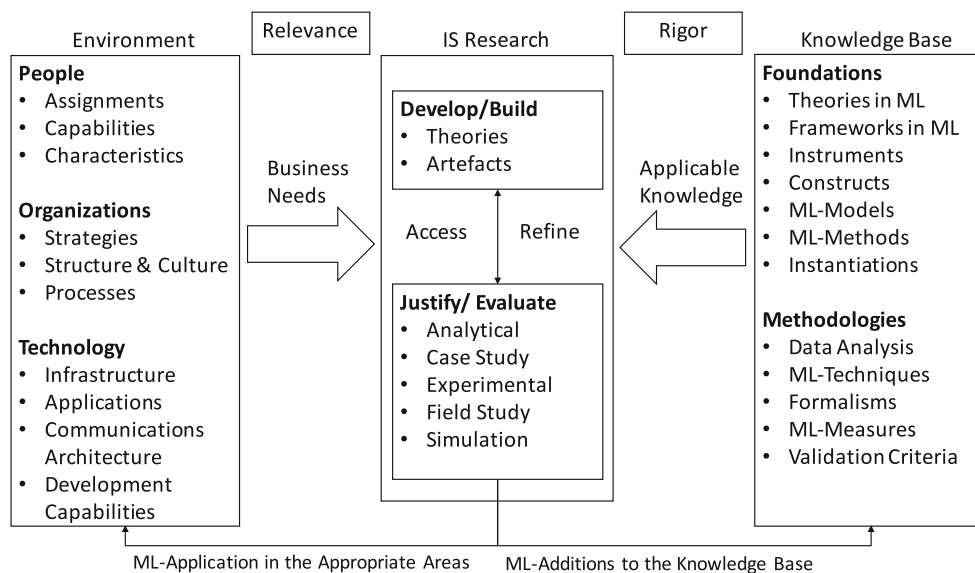


**Fig. 4** Preliminary information systems research framework based on Hevner et al. (2004)

ful enough to observe certain phenomena, it is reasonable to assume, nevertheless, that being able to adjust the behavior of ML-based systems, be it in (field, laboratory) experiments, case studies or action design research, would at the very least drastically increase the external validity of findings. In turn, the findings of IS research may lead to a new impetus for the application of ML in relevant industry domains. In contrast, theoretical aspects that arise from the practical application of certain ML methods may serve as a foundational increase of knowledge for the ML knowledge base. Finally, we believe that by applying ML methods to business problems, IS research theories benefit organizations and society. More extended integration of ML methods in research undertakings seems to be a reasonable step for IS research to maintain the high relevance for theory and practice that it currently enjoys.

## Conclusion

Our bibliographic review and scientometric analysis aim to advance our understanding of the occurrence of research papers concerning ML within the IS field. This paper also aimed to sensitize IS research on applying ML methods based on our bibliographic review in high-level IS journals. We examined two research questions to understand the current state of applied ML research in the field of IS. As a first step to sensitize readers with the problems that are imbued in ML, we provided a brief introduction and overview of AI and ML's related topics, different ML methods, their connections, their advantages, disadvantages, and their limitations. We proceeded with a bibliographic review based on the Senior Scholars Basket of Journals in IS (e.g., Trieu 2017; Khan and Triers 2018) plus the Electronic Markets Journal and Decision Support Systems as proxies for the most important journals in our research domain. We examined 1,919 articles and found 441 articles that make use of ML within six important research databases in the time frame of 2009-2019. However, we only found a small number of articles concerning the application of ML methods in top-ranking IS journals in our bibliographic bibliographic review. We concluded that a lack of trust and expertise in ML methods might be a central reason for the rare application of (sophisticated) ML methods. This conclusion is based on the combination of our bibliographic review, scientometric analysis, and the findings of our nationwide survey of IS scholars in Germany. Of all scholars that participated, 93.22 % have no required ML courses during the PhD, but 74.07 % would like such courses. This detail is particularly relevant, as a large proportion of the survey participants state that they have to work with ML during their PhD studies and

research (share higher to extremely high: 59.6 %, vs. no application: 3.85%), while only 35.48 % of participants had an increased or high share of ML in their studies (Bachelor/Diploma/Master). Due to this lack of in-depth training, a lack of a foundation of trust, confidence, and expertise cannot be ruled out in principle, which promotes uncertainty with methods of higher complexity and lower intuitive interpretability, and may lead to less application of such methods (e.g, Pajares 1996, p. 551). Most of the high-tier IS research journals have contributed only marginally to introducing novel and innovative ML methods to IS research. Based on our results, the Senior Scholars' Basket of Journals in IS appears to have been conservative.

To the best of our knowledge, we are the first who name this problem so clearly. Remarkably, promising calls for papers on AI and ML (e.g., Berente et al. 2019; Jain et al. 2018) have only recently appeared, suggesting that some of the high-tier IS journals have also realized the problem and are willing to tackle it with strengthened efforts.

Overall, ML has progressed dramatically over the last two decades to a widespread commercial key technology (Jordan and Mitchell 2015, p. 255). Disciplines like Computer Science are leading in this research domain. Nevertheless, IS scholars, both proficient or interested in the application of ML, demand an explicit further development of these promising technologies, since they may lead to new insights and allow different analysis approaches besides classic quantitative methods (Sallesa et al. (2017, p. 40); Didimoa et al. (2018, p. 83); Gao et al. (2017, p. 180); Zimbra et al. (2017, p. 108); Abbasi et al. (2010, p. 485)).

Because of the methodological complexity as well as its data-hunger, the choice of ML methods may, for many scholars, not appear as an optimal approach for IS research. As pointed out in our paper, ML methods thrive on large sets of clean and well-structured data for training and testing. This type of data is more easily found in an engineering context and produced by automation and/or sensor-based technology than in a business or a socio-economic context. In this regard and apart from traditional methods and ML, scattered scholars apply methods from the area of Soft Computing (SC) and Computational Intelligence (CI) to socio-economic problems with great success.

CI, as a computational paradigm in the field of ML, has achieved substantial improvements in various fields based on the capability to learning specific tasks from data or experimental observation (Sha et al. 2019, p. 108). As opposed to traditional computing, soft computing deals with approximate models and gives solutions to complex real-life problems. Unlike hard computing, soft computing is tolerant of imprecision, uncertainty, partial truth, and approximations (Ibrahim 2016, p. 34).

Besides being developed and applied in engineering, CI and SC are now growing fast also for management and business purposes, especially within the analytics paradigm, which now is growing strongly and quickly in demand in business and industry (Davenport 2006). The examination of these technologies offers a high business value (Collins et al. 2010, p. 433). Both CI and SC are less sensitive to the quality of the data used for training and testing. The macro-heuristic algorithms are fast and less of a black box nature than more complex ML methods. Future critical research could explore the extent to which CI and SC could improve the limitations in applicability and transparency found for complex ML methods. Nevertheless and similar to the standing of ML in IS research, articles including SC or CI methodology may also only be scarcely found in journals within the Senior Scholars' Basket as the editorial boards appear to be unaware of the potential of these relatively novel and powerful methods.

The results of our investigation warrant further work in this area since the first calls for action to deal with ML in IS research already appeared two decades ago (Wong et al. 1997). Therefore, we call on the IS community to strengthen the efforts concerning the application of ML methods and to conduct thorough research to verify results given by these approaches. Subsequently, we invite fellow scholars and researchers of our domain to tackle the fundamental research questions for applied ML research in IS (Table 4), resulting from our analysis and their reflections. Possibly, a stronger inclusion of ML as a complementary instrument for current research methods - be it case studies, action design research, or experiments - could even have a transformative trend away from monomethodology to an increase in multimethodological studies with more profound insights.

To conclude, our research indicates that IS research is only at the beginning of exploiting and exploring the full potential of ML methods for scientific purposes. We are confident that, by establishing a foundation with verified research results and answering fundamental research questions as a first step, IS researchers and researchers from other economics disciplines will presumably make greater use of ML methods. Ultimately, this could take research in IS and economics to a new level that can create previously unattainable insights.

# Appendix

## A.1 Proprietary algorithm

According to our two-layered approach, in the first step, we use our developed proprietary algorithm to perform a keyword analysis of the publication, which has been retrieved by querying the databases and downloaded as .pdf. The algorithm opens all .pdf files. For each file, the algorithm extracts the text page by page. Each word in the text segment is compared with the previously defined keyword list. Once a match is found, the entire document is tagged accordingly and the algorithm proceeds to the next document. The process will be executed until all PDFs have been checked. Algorithm 1 shows the steps in the pseudo-code. For this purpose, we used the programming language Python in version 3.7 with the PyPDF2 library.

---

**Algorithm 1** Keyword search algorithm.

```
procedure SEARCHFORKEYWORDSINPDF
    LKeywords ← getKeywordList()
    LPapers ← getPaperList()
    for all Paper ∈ LPapers do
        for all Page ∈ Paper do
            text ← getText(Page)
            for all Keyword ∈ LKeywords do
                if text.Contains(Keyword) ==True then
                    tag(Paper)
```

---

## A.2 Questionnaire questions

### Models:

**Which of these methods have you used so far?** Answer option for each model: yes, no unsure, no

**Now, please score the following machine learning methods on a scale from 1 (white box), to 5 (black box)** Answer option for each model: 5-point Likert-Scale, no idea

### Demographics:

**Gender: Please choose only one of the following:** Answer option for each model:female, male, other

**Experience: How long have you been a researcher (in years)?** Answer option: Only numbers may be entered in this field.

**Experience: What is your research position?** Answer option: Prof. Dr., Dr. or PhD Candidate, Research Assistant (without Dr. or PhD Candidate), Student Assistant, Other

**Age: How old are you (in years)?** Answer option: Only numbers may be entered in this field.

**Please choose your highest degree** A nswer option: Associate's degree, Bachelor's degree, Master's degree, Doctoral or PhD degree, Profesional degree, No degree, Other degree.

## Education

**Please choose one of the areas in the context of your highest education.** Choose one of the following answers: Accounting and Finance Biology Computer Science Econometrics, Education, Engineering Information Systems Management (Science), Marketing Mathematics, Medical, Logistics Statistics, Other

**Relative orientation of the course of study in the context of highest degree. If you have a master's degree then refer it to the master's degree. If you have a diploma then refer it to the diploma. Please estimate the share of the respective sub-areas of the course of study.** For: Relative statistics, relative mathematics, relative Computer Science, relative data science, relative machine learning

Answer option: 7-point Likert-Scale, no share at all.

**Did you have any (obligatory) courses in the process of your PhD studies in the field of Machine Learning?** Answer options: yes, no

## Personal opinion

**Would you recommend or wish for (obligatory) courses on Machine Learning during the PhD program in your field?** Answer options: yes, no

**How relevant is ML in IS Research? Please choose one option, from absolutely not relevant to absolutely relevant, for each questions.** Answer option: 7-point Likert-Scale, no share at all.

### A.3 Questionnaire detail view of results

The values reflect the percentage of votes for each model in relation to the black box rating for each row.

Table 8 shows ML methods sorted according to their transparency (K-Medians = White box; ANN = Black box) based on the assessment of the IS researchers.

**Fig. 5** Normalized and sorted results for: Now, please score the following machine learning methods on a scale from 1 = White box, to 5 = Black box
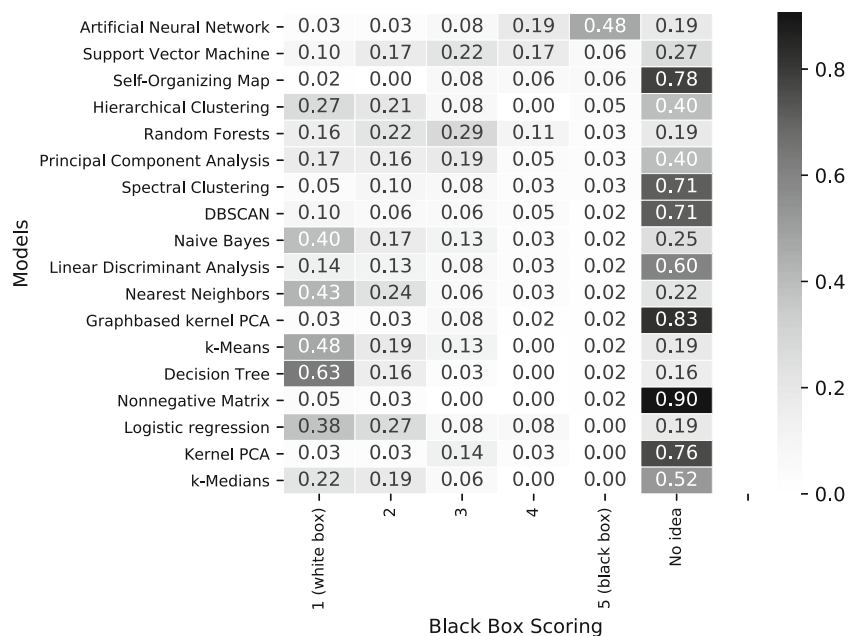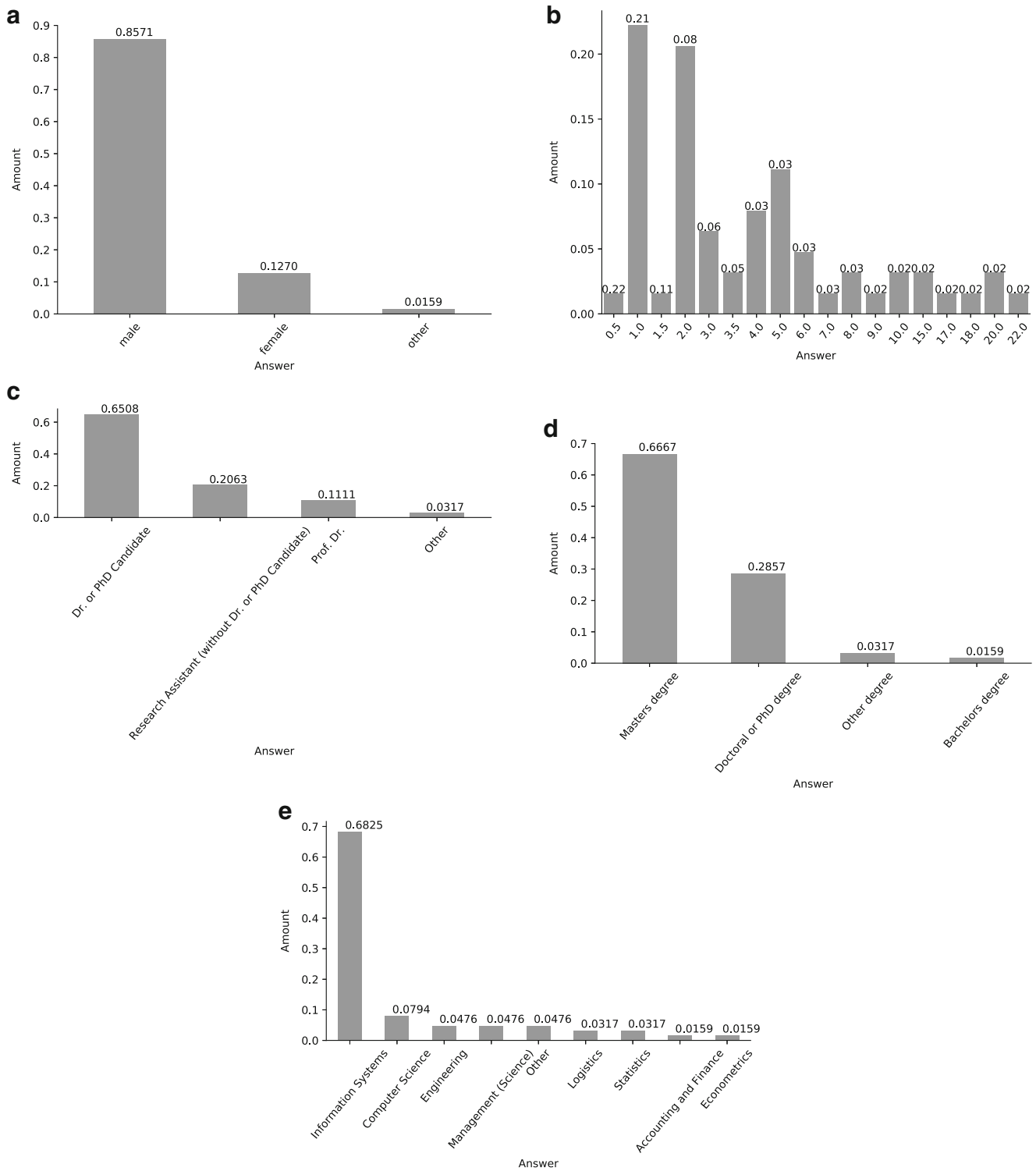
**Fig. 6** Demographics and education. **a** Gender. **b** Experience: How long have you been a researcher (in years)? **c** Experience: What is your research position? **d** Please choose your highest degree. **e** Please choose one of the areas in the context of your highest education
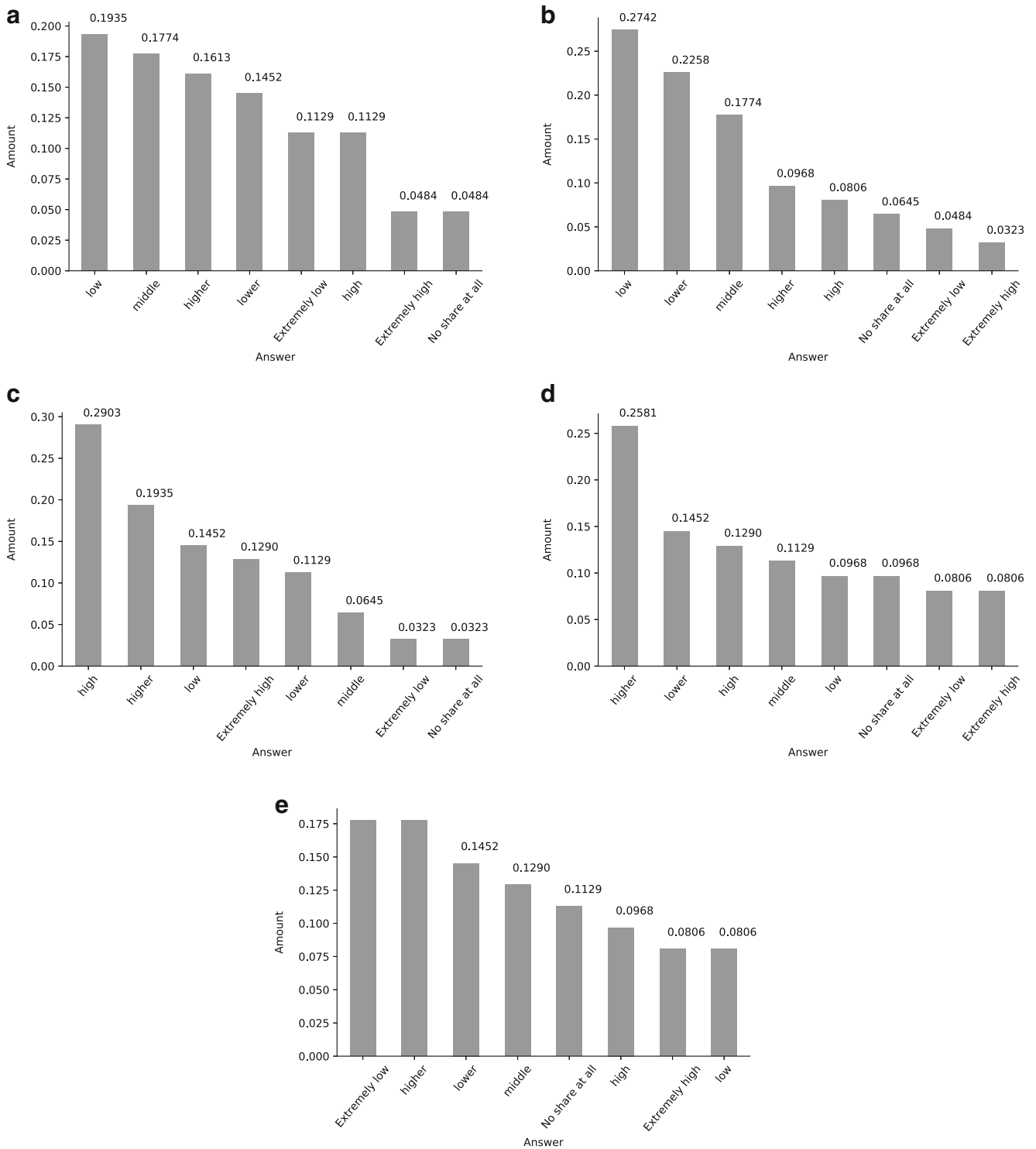
**a**



**b**



**c**



**d**



**e**



**Fig. 7** Share of the respective sub-areas of the course of study. **a** Please estimate the share of the respective sub-areas of the course of study (Relative Statistics). **b** (Relative mathematics). **c** (Relative computer science). **d** (Relative data science). **e** (Relative machine learning)
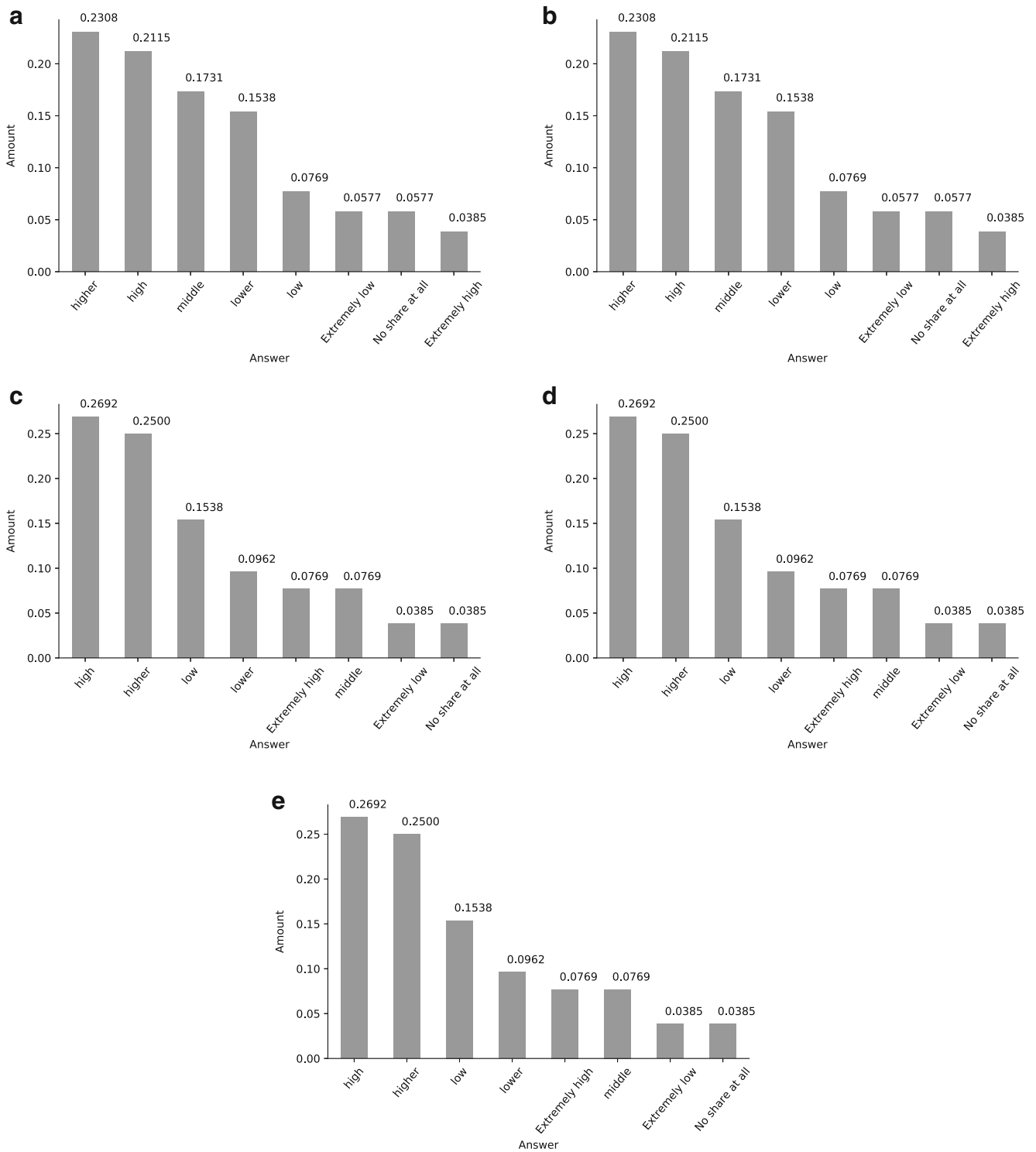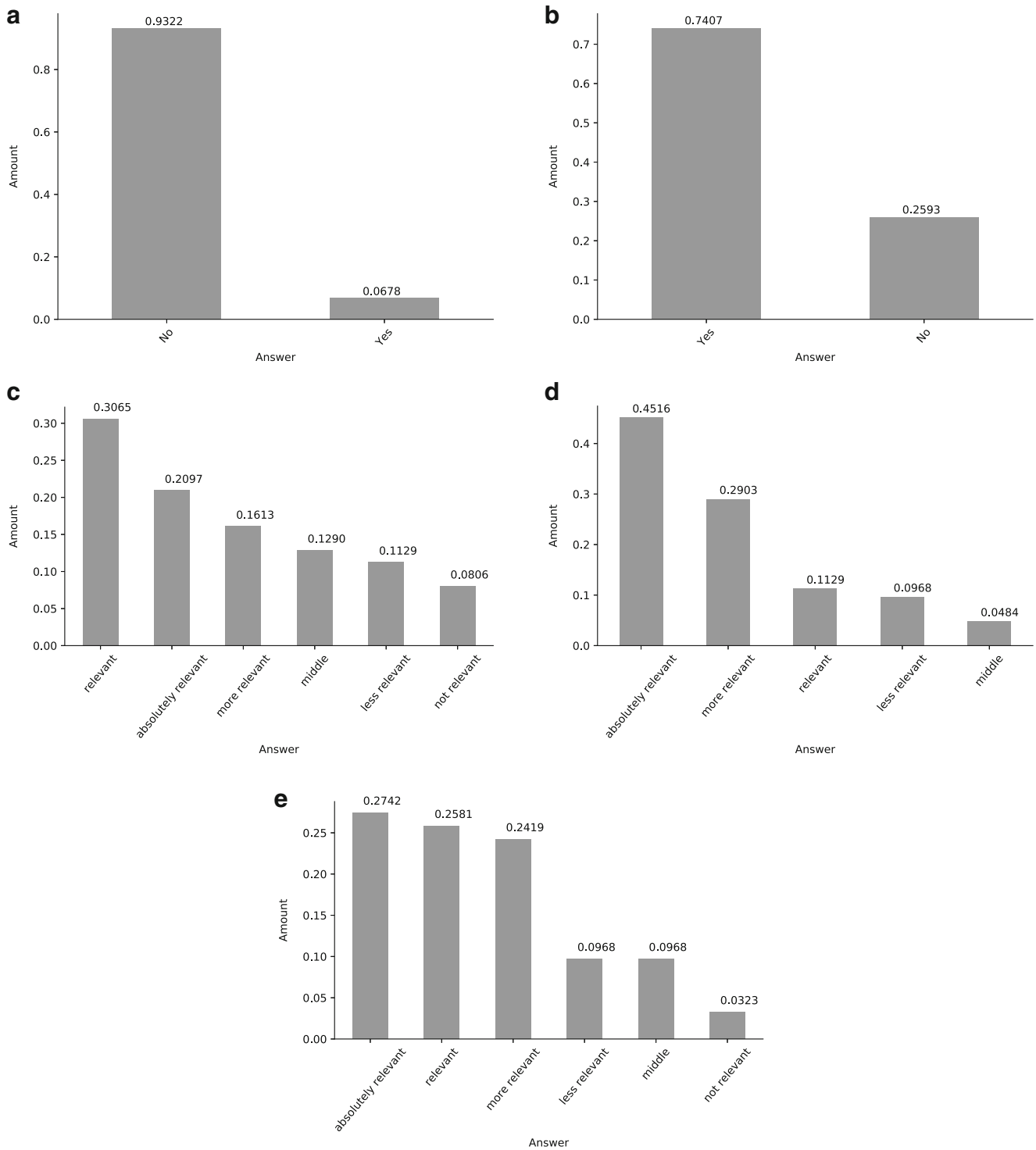
**Fig. 8** In the process of your PhD studies, please estimate the share of the respective sub-areas during your PhD studies. **a** Please estimate the share of the repective sub-areas during your PhD studies (Relative statistics). **b** (Relative mathematics). **c** (Relative computer science). **d** (Relative data science). **e** (Relative machine learning)

**Fig. 9** Relevance of ML in the PhD process and for the IS community. **a** Did you have any (obligatory) courses in the process of your PhD studies in the field of ML? **b** Would you recommend or wish for (obligatory) courses on ML during the PhD programm in your field. **c** How important is ML in IS currently? **d** How important is ML in the future? **e** How important do you think ML is for the progress of the IS discipline?
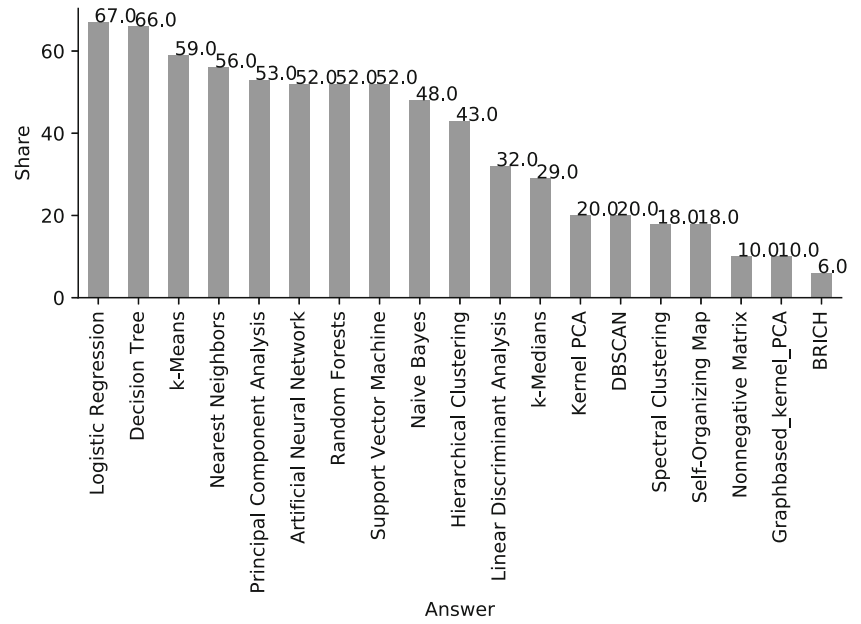
**Fig. 10** Used models



**Table 8** Usage of ML Methods in IS Publications. The methods are ordered according to their degree of the black box characteristics (top = White box method, bottom = Black box method)

|                              | EJIS | ISJ  | ISR  | JAIS | JIT  | JMIS | JSIS | MISQ | ELMA | DSS  |
|------------------------------|------|------|------|------|------|------|------|------|------|------|
| K-Medians                    | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.67 |
| Kernel PCA                   | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.67 |
| Logistic regression          | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.17 | 0.33 |
| Nonnegative Matrix           | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 |
| Decision Tree                | 0.00 | 0.19 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.04 | 0.01 | 0.73 |
| k-Means                      | 0.00 | 0.35 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.15 | 0.46 |
| Graph-based Kernel PCA       | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Nearest Neighbors            | 0.00 | 0.33 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.64 |
| Linear Discriminant Analysis | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.17 | 0.33 |
| Naïve Bayes                  | 0.00 | 0.16 | 0.05 | 0.02 | 0.00 | 0.02 | 0.00 | 0.05 | 0.03 | 0.67 |
| DBSCAN                       | 0.00 | 0.38 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.38 |
| Spectral Clustering          | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.67 |
| Principal Component Analysis | 0.00 | 0.18 | 0.04 | 0.00 | 0.00 | 0.00 | 0.18 | 0.04 | 0.04 | 0.54 |
| Random Forests               | 0.02 | 0.12 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.06 | 0.76 |
| Hierarchical Clustering      | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 | 0.00 |
| Self-Organizing Map          | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Support Vector Machine       | 0.00 | 0.12 | 0.01 | 0.01 | 0.00 | 0.01 | 0.00 | 0.02 | 0.05 | 0.79 |
| Artificial Neural Network    | 0.00 | 0.07 | 0.02 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.05 | 0.83 |

# References

Abbasi, A., Sarker, S., Chiang, R.H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), 1–33.

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., Nunamaker, J.F. (2010). Detecting fake websites: The contribution of statistical learning theory. *MIS Quarterly*, *34*(4), 435–461.

Agarwal, R., & Dhar, V. (2014). Editorial - big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, *25*(3), 443–448.

Aleksander, I. (2017). Partners of humans: a realistic assessment of the role of robots in the foreseeable future. *Journal of Information Technology*, *32*(1), 1–9.

Altman, N.S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175–185.

Arazy, O., Daxenberger, J., Lifshitz-Assaf, H., Nov, O., Gurevych, I. (2016). Turbulent stability of emergent roles: The dualistic nature of Self-Organizing knowledge coproduction. *Information Systems Research*, *27*(4), 792–812.

Au, T.C. (2018). Random forests, decision trees, and categorical predictors: The "Absent" levels problem. *The Journal of Machine Learning Research*, *19*(1), 1737–1766.

Bach, F. (2017). Breaking the curse of dimensionality with convex neural networks. *The Journal of Machine Learning Research*, *18*(1), 629–681.

Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J., Zhao, J.L. (2016). Transformational issues of big data and analytics in networked business. *MIS Quarterly*, *40*(4), 807–018.

Baier, L., Jöhren, F., Seebacher, S. (2019). Challenges in the deployment and operation of machine learning in practise. *Proceedings of the 27th European Conference on Information Systems (ECIS)*, *8*(14), 1–15.

Basden, A. (2010). On using spheres of meaning to define and dignify the IS discipline. *International Journal of Information Management*, *30*(1), 13–20.

Basti, E., Kuzey, C., Delen, D. (2015). Analyzing initial public offerings short-term performance using decision trees and SVMs. *Decision Support Systems*, *73*(5), 15–27.

Bélanger, F., & Crossler, R.E. (2011). Privacy in the digital age: a review of information privacy research in information systems. *MIS Quarterly*, *35*(4), 217–235.

Belson, W.A. (1959). Matching and prediction on the principle of biological classification. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *8*(4), 65–75.

Benbasat, I., Goldstein, D.K., Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly*, *11*(3), 369–386.

Benbasat, I., & Zmud, R.W. (1999). Empirical research in information systems: The practice of relevance. *MIS Quarterly*, *23*(2), 3–16.

Benbasat, I., & Zmud, R.W. (2003). The identity crisis within the is discipline: Defining and communicating the discipline's core properties. *MIS Quarterly*, *27*(2), 183–194.

Benbernou, S., & Ouziri, M. (2017). Enhancing data quality by cleaning inconsistent big RDF data. *IEEE International Conference on Big Data*, *11*(14), 74–79.

Berente, N., Gu, B., Recker, J., Santhanam, R. (2019). Managing AI. Call for Papers, MIS Quarterly.

Bishop, C.M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Bose, I., & Mahapatra, R.K. (2001). Business data mining a machine learning perspective. *Information & Management*, *39*(12), 211–225.

Boudreau, M.-C., Gefen, D., Straub, D.W. (2001). Validation in information systems research: a state-of-the-art assessment. *MIS Quarterly*, *25*(1), 1–16.

Casey, B., Farhangi, A., Vogl, R. (2019). Rethinking explainable machines: the GDPR's right to explanation debate and the rise of algorithmic audits in enterprise. *Berkeley Tech. LJ*, *34*, 143.

Chen, H., Chiang, R.H.L., Storey, V.C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188.

Chen, W., & Hirschheim, R. (2014). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal*, *14*(1), 197–235.

Chollet, F. (2018). *Deep learning with R*. Shelter Island: Manning Publications.

Chu, H., & Ke, Q. (2017). Research Methods: What's in the Name? *Library and Information Science Research*, *39*(1), 284–294.

Cnudde, S.D., & Martens, D. (2015). Loyal to your city? a data mining analysis of a public service loyalty program. *Decision Support Systems*, *73*(5), 74–84.

Collins, J., Ketter, W., Gini, M. (2010). Flexible decision support in dynamic Inter-Organisational networks. *European Journal of Information Systems*, *19*(4), 436–448.

Cresci, S., Pietro, R.D., Petrocchi, M., Spognardi, A., Tesconi, M. (2015). Fame for sale: Efficient detection of fake twitter followers. *Decision Support Systems*, *80*(12), 56–71.

Culnan, M.J., & Williams, C.C. (2011). How ethics can enhance organizational privacy: Lessons from the choicepoint and TJX data breaches. *MIS Quarterly*, *33*(4), 673–687.

Davenport, T.H. (2006). Competing on analytics. *Harvard Business Review*, *1*(6), 1–10.

Davison, R.M. (2017). Transition arrangements to a new editorial structure. *Information Systems Journal*, *27*(1), 1–3.

Davison, R.M., Powell, P., Trauth, E.M. (2012). ISJ inaugural editorial. *Information Systems Journal*, *22*(4), 257–260.

Delen, D., Zaim, H., Kuzey, C., Zaim, S. (2013). A comparative analysis of machine learning systems for measuring the impact of knowledge management practices. *Decision Support Systems*, *54*(2), 1150–1160.

Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research - best of the web. *IEEE Signal Processing Magazine*, *6*(29), 141–142.

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64–73.

Didimoa, W., Giamminonnia, L., Liotta, G., Montecchiania, F., Pagliuca, D. (2018). A visual analytics system to support tax evasion discovery. *Decision Support Systems*, *110*(5), 71–83.

Evermann, J., Rehseb, J.-R., Fettkeb, P. (2017). Predicting process behaviour using deep learning. *Decision Support Systems*, *100*(8), 129–140.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37–54.

Feldman, M., Friedler, S.A., Moeller, J., Scheidegger, C., Venkatasubramanian, S. (2015). Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International conference on knowledge discovery and data mining* (pp. 259–268).

Fu, X., Chen, X., Shi, Y.-T., Bose, I., Cai, S. (2017). User segmentation for retention management in online social games. *Decision Support Systems*, *101*(9), 51–68.

Gao, Y., Xu, A., Hu, P.J.-H., Cheng, T.-H. (2017). Incorporating association rule networks in feature Category-Weighted naive bayes model to support weaning decision making. *Decision Support Systems*, *96*(4), 27–38.

Gerber, S., & Horenko, I. (2015). Improving clustering by imposing network information. *Science Advances*, *1*(7), 1–8.

Goes, P.B. (2014). Editor's comments: Big Data and IS Research. MIS Quarterly.

Gong, J., Abhisek, V., Li, B. (2018). Examining the impact of keyword ambiguity on search advertising performance: a topic model approach. MIS Quarterly, In Press.

Goodhue, D.L., Lewis, W., Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly*, *36*(3), 981–1001.

Goodhue, D.L., Lewis, W., Thompson, R. (2017). A multicollinearity and measurement error statistical blind spot: Correcting for excessive false positives in regression and PLS. *MIS Quarterly*, *3*(41), 667–684.

Hevner, A.R., March, S.T., Park, J., Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K. (2016). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*(12), 489–501.

Hutson, M. (2017). AI glossary: artificial intelligence, in so many words. *Science*, *357*(6346), 19.

Ibrahim, D. (2016). An overview of soft computing. *12th International conference on application of fuzzy systems and soft computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria*, *102*(1), 34–38.

Jain, H., Padmanabhan, B., Pavlou, P.A., Santanamd, R.T. (2018). Special issue of information systems Research—Humans, algorithms, and augmented intelligence: The future of work, organizations, and society. *Information Systems Research*, *29*(1), 250–251.

Jordan, M.I., & Mitchell, T.M. (2015). Machine learning - trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Junior, M.R., & da Silva, F.J. (2017). An on-line algorithm for cluster detection of mobile nodes through complex event processing. *Information Systems*, *64*(3), 303–320.

Kamiński, B., Jakubczyk, M., Szufel, P. (2018). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*, *26*(1), 135–159.

Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U. (1999). When is nearest neighbor meaningful? International conference on database theory, pp 217–235.

Khan, G.F., & Triers, M. (2018). Assessing the long-term fragmentation of information systems research with a longitudinal multi-network analysis. *European Journal of Information Systems*, *28*(4), 370–393.

Khan, G.F., & Wood, J. (2016). Knowledge networks of the information technology management domain: a social network analysis approach. *Communications of the Association for Information Systems*, *39*(18), 367–397.

Kitchenham, B. (2004). Procedures for performing systematic reviews. Tech. Rep NICTA technical report 0400011T.1, Empirical Software Engineering National ICT Australia Ltd., Australia.

Kohli, R., & Melville, N.P. (2018). Digital innovation: a review and synthesis. *Information Systems Research*, *28*(4), 1–24.

Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Reynolds, J., Melnikov, A., Lunova, N., Reblitz-Richardson, O. (2019). PyTorch Captum. https://github.com/pytorch/captum.

Krauss, C., Do, X.A., Huck, N. (2017). Deep neural networks, Gradient-Boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689–702.

Lara, J.A., Lizcano, D., Martínez, A., Pazos, J. (2014). Data preparation for KDD through automatic reasoning based on description logic. *Information Systems*, *44*(8), 54–72.

Lau, R.Y.K., Liao, S.S.Y., Wong, K.F., Chiu, D.K.W. (2012). Web 2.0 environmental scanning and adaptive decision support for business mergers and acquisitions. *MIS Quarterly*, *36*(4), 1239–1268.

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, *521*(1), 436–444.

Lee-Post, A., & Pakath, R. (2019). Numerical, secondary big data quality issues, quality threshold establishment, & guidelines for journal policy development. *Decision Support Systems*, *126*(1), 113–135.

Lehmann, R., & Wohlrabe, K. (2017). Who is the 'Journal grand master'? a new ranking based on the elo rating system. *Journal of Informetrics*, *11*(3), 800–809.

Lewis, B.R., Templeton, G.F., Luo, X. (2007). A scientometric investigation into the validity of IS. *Journal Quality Measures. Journal of the Association for Information Systems*, *8*(12), 35.

Leydesdorff, L. (2001). The challenge of scientometrics: the development, measurement, and self-organization of scientific communications. Universal-Publishers.

Li, J., Tso, K.F., Liu, F. (2017). Profit earning and monetary loss bidding in online entertainment shopping: The impacts of bidding patterns and characteristics. *Electronic Markets*, *25*(10), 77–90.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.

Lorbeer, B., Kosareva, A., Deva, B., Softic, D., Ruppel, P., Küpper, A. (2018). Variations on the clustering algorithm BIRCH. *Big Data Research*, *11*(3), 44–53.

Lowry, P.B., Gaskin, J., Humpherys, S.L., Moody, G.D., Galletta, D.F., Barlow, J.B., Wilson, D.W. (2013). Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: do expert journal assessments add value? *MIS Quarterly*, *37*(4), 993–1012.

Lowry, P.B., Romans, D., Curtis, A.M. (2004). Global journal prestige and supporting disciplines: a scientometric study of information systems journals. *Journal of the Association for Information Systems*, *5*(2), 29–80.

Lozano, M.G., Schreiber, J., Brynielsson, J. (2017). Tracking geographical locations using a Geo-Aware topic model for analyzing social media data. *Decision Support Systems*, *99*(1), 18–29.

Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (pp. 4765–4774).

Maass, W., Parsons, J., Purao, S., Storey, V.C., Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems*, *19*(12), 1253–1273.

Marban, O., Segovia, J., Menasalvas, E., Fernandez-Baizan, C. (2009). Toward data mining engineering: a software engineering approach. *Information Systems*, *34*(1), 87–107.

Marcoulides, G.A., Chin, W.W., Saunders, C. (2009). A critical look at partial least squares modeling. *MIS Quarterly*, *1*(33), 667–684.

Martens, D., & Provost, F. (2014). Explaining data-driven document classifications. *MIS Quarterly*, *38*(1), 73–99.

Martens, D., Provost, F., Clark, J., de Fortuny, E.J. (2016). Mining massive Fine-Grained behavior data to improve predictive analytics. *MIS Quarterly*, *40*(4), 869–888.

Mason, R.O. (1986). Four ethical issues of the information age. *MIS Quarterly*, *10*(1), 5–12.

Meyer, G., Adomavicius, G., Johnson, P.E., Elidrisi, M., Rush, W.A., Sperl-Hillen, J.M., O'Connor, P.J. (2018). A machine learning approach to improving dynamic decision making. *Information Systems Research*, *25*(2), 239–263.

Mingers, J., & Yang, L. (2017). Evaluating journal quality: a review of journal citation indicators and ranking in business and management. *European Journal of Operational Research*, *257*(1), 323–337.

Mjolsness, E., & DeCoste, D. (2001). Machine learning for science: State of the art and future prospects. *Science*, *293*(5537), 2051–2055.

Müller, O., Junglas, I., vom Brocke, J., Debortoli, S. (2016). Utilizing big data analytics for information systems research: challenges, Promises and Guidelines. *European Journal of Information Systems*, *25*(5), 289–302.

Nwankpa, J.K., & Datta, P. (2017). Balancing exploration and exploitation of IT resources: The influence of digital business intensity on perceived organizational performance. *European Journal of Information Systems*, *26*(5), 469–488.

Olden, J.D., & Jackson, D.A. (2002). Illuminating the "Black Box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*(1-2), 135–150.

Pai, H.-T., Wu, F., Hsueh, P.-Y.S. (2014). A relative patterns discovery for enhancing outlier detection in categorical data. *Decision Support Systems*, *67*(11), 90–99.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578.

Pavlou, P.A. (2011). State of the information privacy literature: where are we now and where should we go? *MIS Quarterly*, *35*(4), 977–988.

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. New York City: Basic Books.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572.

Pinto, T., Barreto, J., Praca, I., Sousa, T.M., Vale, Z., Pires, E.S. (2015). Six thinking hats: a novel metalearner for intelligent decision support in electricity markets. *Decision Support Systems*, *79*(11), 1–11.

Piri, S., Delen, D., Liu, T. (2018). A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets. *Decision Support Systems*, *106*(2), 15–29.

Puschmann, T. (2017). Fintech. *Business & Information Systems Engineering*, *59*(1), 69–76.

Rai, A. (2020). Explainable AI: From Black box to Glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141.

Recker, J. (2013). *Scientific research in information systems - a beginner's guide Berlin*. Heidelberg: Springer.

Reicha, Y., & Barai, S. (1999). Evaluating machine learning models for engineering problems. *Artificial Intelligence in Engineering*, *13*(1), 257–272.

Ribeiro, M.T., Singh, S., Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining* (pp. 1135–1144).

Rittgen, P. (2009). Self-organization of interorganizational process design. *Electronic Markets*, *19*(10), 189–199.

Rojasa, W.A.C., & Villegas, C.J.M. (2013). Graphical representation and exploratory visualization for decision trees in the KDD process. *Procedia - Social and Behavioral Sciences*, *73*(2), 136–144.

Roweis, S.T., & Saul, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*(5500), 2323–2326.

Russell, S.J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia: Pearson.

Sallesa, T., Rochab, L., Mouraob, F., Goncalvesa, M., Viegasa, F., Meira, W. Jr.. (2017). A two-stage machine learning approach for temporally-robust text classification. *Information Systems*, *69*(9), 40–58.

Scholz, M., Franz, M., Hinz, O. (2016). The ambiguous identifier clustering technique. *Electronic Markets*, *26*(2), 143–156.

See-To, E.W.K., & Yang, Y. (2017). Market sentiment dispersion and its effects on stock return and volatility. *Electronic Markets*, *27*(4), 283–296.

Seiler, C., & Wohlrabe, K. (2014). How robust are journal rankings based on the impact factor? Evidence from the economic sciences. *Journal of Informetrics*, *8*(4), 904–911.

Sha, H., Xua, P., Yanga, Z., Chena, Y., Tang, J. (2019). Overview of computational intelligence for building energy system design. *Renewable and Sustainable Energy Reviews*, *108*(1), 76–90.

Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, *12*(1), 217–222.

Shahrivari, S., & Jalili, S. (2016). Single-pass and linear-time k-means clustering based on MapReduce. *Information Systems*, *60*(9), 1–12.

Siering, M., Deokar, A.V., Janze, C. (2018). Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decision Support Systems*, *107*(3), 62–63.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(10), 354–359.

Singh, A., & Tucker, C.S. (2017). A machine learning approach to product review disambiguation based on function, form and behavior classification. *Decision Support Systems*, *97*(5), 81–91.

Snoek, J., Larochelle, H., Adams, R.P. (2012). Practical Bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems* (pp. 2951–2959).

Sokolova, K., Perez, C., Lemerciera, M. (2017). Android application classification and anomaly detection with graph-based permission patterns. *Decision Support Systems*, *93*(1), 62–76.

Straub, D. (2006). The value of scientometric studies: an introduction to a debate on IS as a reference discipline. *Journal of the Association for Information Systems*, *7*(5), 241.

Straub, D.W. (1989). Validating instruments in MIS research. *MIS Quarterly*, *13*(2), 147–169.

Swiderski, B., Kurek, J., Osowski, S. (2012). Multistage classification by using logistic regression and neural networks for assessment of financial condition of company. *Decision Support Systems*, *52*(2), 539–547.

Topuz, K., Zengul, F.D., Dag, A., Almehmi, A., Yildirim, M.B. (2018). Predicting graft survival among kidney transplant recipients: a bayesian decision support model. *Decision Support Systems*, *106*(2), 97–109.

Trieu, V.-H. (2017). Getting value from business intelligence systems: a review and research agenda. *Decision Support Systems*, *93*(1), 111–124.

Valle, M.A., Ruz, G.A., Masías, V.H. (2017). Using self-organizing maps to model turnover of sales agents in a call center. *Applied Soft Computing*, *60*(11), 763–774.

van Riessen, B., Negenborn, R.R., Dekker, R. (2016). Real-time container transport planning with decision trees based on offline obtained optimal solutions. *Decision Support Systems*, *89*(9), 1–16.

Verhulst, P.-F. (1883). Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathématique et Physique*, *10*(1), 113–121.

Vidgen, R., Henneberg, S., Naude, P. (2007). What sort of community is the European conference on information systems? a social network analysis 1993–2005. *European Journal of Information Systems*, *16*(1), 5–19.

Walczak, S., & Velanovich, V. (2018). Improving prognosis and reducing decision regret for pancreatic cancer treatment using artifiial neural networks. *Decision Support Systems*, *106*(2), 110–118.

Wan, Y. (2015). The Matthew effect in social commerce. *Electronic Markets*, *25*(3), 33–324.

Wang, G.A., Atabakhsh, H., Chen, H. (2011). A hierarchical naïve Bayes model for approximate identity matching. *Decision Support Systems*, *51*(3), 413–423.

Wang, L., Gopal, R., Shankar, R., Pancras, J. (2015). On the brink: predicting business failure with mobile location-based checkins. *Decision Support Systems*, *76*(8), 3–13.

Wayne, G., & Pasternack, A. (2011). Canny minds and uncanny questions. *Science*, *333*(6047), 1223–1224.

Webster, J., & Watson, R.T. (2002). Analyzing the past to prepare for the future: writing a literature review. *MIS Quarterly*, *26*(2), xiii–xxiii.

Wolfswinkel, J.F., Furtmueller, E., Wilderom, C.P. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, *22*(1), 45–55.

Wong, B.K., Bodnovich, T.A., Selvi, Y. (1997). Neural network applications in business: a review and analysis of the literature (1988–1995). *Decision Support Systems*, *19*(4), 301–320.

Yang, M., Adomavicius, G., Burtch, G., Ren, Y. (2018). Mind the gap: Accounting for measurement error and misclassification in variables generated via data mining. *Information Systems Research*, *29*(1), 4–24.

Zhang, L., Mistry, K., Lim, C.P., Neoh, S.C. (2017). Feature selection using firefly optimization for classification and regression models. *Decision Support Systems*, *106*(2), 64–85.

Zimbra, D., Sarangee, K.R., Jindal, R.P. (2017). Movie aspects, tweet metrics, and movie revenues: the influence of iOS vs Android. *Decision Support Systems*, *102*(10), 98–109.