Applied Data Science and Analysis
Vol.2023, **pp**. 118-125
DOI: https://doi.org/10.58496/ADSA/2023/011 ; ISSN: 3005-317X
https://mc04.manuscriptcentral.com/adsa

Review Article

Challenges in AutoML and Declarative Studies Using Systematic Literature Review

Eman Thabet Khalid [1], *, Abdulla J. Y. Aldarwish [1], Ali A.Yassin[1]

[1] *Department of Computer Sciences, Education College for Pure Sciences, University of Basrah, 6100, Iraq*

ARTICLE INFO

Article History

Received 04 Aug 2023
Accepted 19 Oct 2023
Published 15 Nov 2023

Keywords

Machine learning

AI

Declarative ML

AutoML

ML

ABSTRACT

Machine Learning (ML) technologies have become essential tools, transforming industries and unlocking incredible potential in various fields. ML is now widely used for data-driven decision-making and predictive analytics across fields like healthcare, finance, transportation, and more. However, building and implementing ML models can be complex and time-consuming, often requiring programming proficiency and data science skills. Despite significant progress in ML, non-experts often struggle with selecting algorithms, optimizing models, and deploying ML solutions. This paper conducts a systematic literature review to explore challenges in the area of machine learning based on multiple categories involving features engineering and data extraction, learning model structure and activities, learning-based analysis and visualization, analysis algorithms in data-based systems, machine learning algorithms and systems development, and declarative ML-based prediction. Addressing these challenges underlines the importance of following AutoML and Declarative ML strategies in simplifying the ML process.

1. INTRODUCTION

Machine Learning (ML) technologies have become crucial in changing industries and unlocking incredible potential in different areas. The use of ML has brought about a new era of data-driven decision-making and predictive analytics, impacting industries like healthcare, finance, transportation, and more. ML algorithms have the ability to recognize patterns, learn from data, and make informed predictions, speeding up operations and offering insights that were once unimaginable.

Unfortunately, existing Artificial Intelligence (AI) and machine learning (ML) technologies are not adequately flexible. They do not deliver easy means for designing applications for areas of experts who are not experts in AI; instead, they offer complex solutions across numerous dimensions. Furthermore, even for AI experts wanting to evaluate their novel ideas and algorithms, extensive experimentation and evaluation are necessary. This is because AI has reached the severe phase of foundation research application to issues of the real world. Large AI systems creation requires substantial program writing abilities, as well as the ability to deal with a variety of thinking and knowledge paradigms and methodologies at a relatively low range of declarative. Because theoretic understanding could be utilized to automatically abstract these details, it also demands serious investigation and experiments for selecting the appropriate model, feature selection, and adjusting parameters, Yet it is short and small [1].

Typical languages of programming and software engineering methodologies were not created to address the issues faced by users of AI systems, such as coping with cluttered, actual data at the appropriate level of declarative. As a result, creative schemes are needed to smoothly allow embedded trainable models while excluding the utmost details of low-limit and facilitating cognitive with respect to them at the appropriate degree of declarative and abstraction. There is a necessity to enhance current frameworks and their abilities for programming-based learning and complex design of AI systems, such as:- Simple interaction with untreated, diverse data, high-limit and natural abstractions for identifying requirements of application, obvious means for describing domain-data knowledge and expressing doubts, reaches to a variety of learning, and cognitive techniques, prediction, and the talent to re-utilize, chain, and merge models, as well as execute flexible inference on complicated models/pipelines [2].

In the same manner, large-scale machine learning (ML) makes use of sophisticated analytics to mine massive datasets for intriguing patterns and build reliable prediction models. Although conventional frameworks, and tools such as Matlab, Weka, SAS, SPSS, or R offer rich capability, they find it difficult to provide scalable analytics, with the exception of

*Corresponding author. Email: eman.alasadi@uobasrah.edu.iq

specialized programs. Due to the data-intensive nature, paralleling frameworks of data such as Flink [3], Spark[4], and MapReduce[5] are being utilized more frequently for economical parallelization on commodity hardware. On the other hand, large-scale computing naturally makes ML algorithm specification more challenging, particularly in terms of scalable and effective execution [6]. The most common tools for ML of large-scale tasks nowadays are ML libraries of large-scale type such as MLlib (aka SparkML) [7], Mahout [8], and MADlib[9, 10]. These libraries offer algorithms with predefined dispersed runtime schedules and frequently reveal the representation of physical data that lies beneath them. Even while these libraries are extremely useful tools for end users, writing new algorithms or modifying old ones requires a lot of work because it necessitates understanding algorithms of ML, dispersed implementation, and architecture of the paralleling data underlying. In a similar vein, enhancements frequently necessitate adjusting each and every algorithm to utilize them fully[6].

However, despite tremendous advances in machine learning, end-users face a complex challenge in realizing its full potential. Non-experts frequently struggle with the complexities of algorithmic selection, model optimization, and the overall complexity of deploying machine learning solutions. These challenges aim to overcome the main obstacles end users encounter while implementing ML. Technical complexity is one such obstacle, as conventional machine learning workflows necessitate a high level of technical proficiency, particularly in programming and statistics. This complexity keeps many people and organizations from utilizing machine learning. Another issue involves a time-consuming process; developing and implementing machine learning models can be a long, iterative process that frequently calls for constant trial and error. The use of ML for time-sensitive applications may be constrained by this time commitment. A third issue includes limited transparency; users may find it challenging to comprehend how models arrive at their predictions due to the complexity of ML algorithms. This lack of transparency might erode confidence in ML-driven judgments by raising issues with bias and interpretability. Lack of control is another issue; traditional machine learning algorithms frequently limit user control over the model-building process, making it difficult to add relevant domain expertise or adapt models for specific use cases [1-6].

Two closely related strategies in machine learning have emerged to address this problem, Automated Machine Learning (AutoML), and Declarative Machine Learning (Declarative ML), offer distinct viewpoints on the machine learning workflow. Both AutoML and Declarative ML aim to simplify the process of generating machine learning models, yet they differ in their core concepts and interactions with users. The primary distinction lies in the fact that AutoML focuses on automating the entire process of developing machine learning models, including the engineering of features, selection of models, tuning hyperparameters, and model deployment. The ultimate goal is to make machine learning more accessible, particularly to individuals who are not experts in the field[11]. In contrast, Declarative ML centres on providing a high-level abstraction. Instead of specifying how the model should be implemented, it emphasizes the declaration of desired outcomes, allowing users to express their intent without delving into the specifics of low-level implementation details. Automation is a pivotal component of AutoML, as the system independently seeks and selects the best model topologies, hyperparameters, and other configurations according to predetermined standards. Declarative ML, on the other hand, underscores abstraction and provides users with the ability to explicitly state their goals, limitations, and intended results, leaving the system to determine the best approach to achieve those objectives [12].

Due to its intended audience of non-experts, AutoML often minimizes the need for user intervention. Typically, users input the task and data, and the system takes care of the rest. Conversely, Declarative ML offers a greater degree of freedom, enabling users to guide the learning process by expressing their preferences, limitations, and domain-specific expertise. Given that AutoML is typically more automatic and less flexible, users may find it challenging to fine-tune certain algorithmic decisions or specifics. In contrast, Declarative ML allows users to exert control over the learning process, drawing on their domain knowledge and expertise and provides more flexibility and customization choices. The primary objective of AutoML is to make machine learning processes more accessible and easy by automating them, enabling a wider range of individuals to familiarize themselves with machine learning concepts. On the other hand, Declarative ML achieves simplicity through high-level abstractions, allowing users to express their intents more naturally [11, 13].

This paper aims to explore the previous research in AI and machine learning, seeking to understand the challenges these technologies face. Specifically, by exploring ML problems focusing on the ways to blend old ideas into new approaches called Declarative ML and AutoML. This study's contribution is to identify existing challenges in machine learning, aligning with the emerging concepts of AutoML and Declarative ML that can simplify ML technologies and make them more effective for end users of AI.

The following is the structure of this document. Section 2 describes the approach for undertaking a systematic literature review. In accordance with the AutoML and Declarative ML studies, Section 3 highlights the challenges in ML technologies across various categories. Finally, Section 4 concludes this contribution.

2. THE REVIEW APPROACH

Several bibliographic citation databases were employed in this analysis to conduct a systematic review. We explored five widely used digital databases, namely IEEE Xplore, ACM Digital Library, Web of Science (WoS), Science Direct, and Scopus, to identify relevant publications. IEEE Xplore offers comprehensive papers and summaries in the fields of computer science, electronics, and electrical engineering, covering a wide range of technical and scientific publications. Science Direct is a reputable source for publications in science, technology, and engineering. The ACM Digital Library encompasses academic research publications across various fields. Scopus includes reliable resources from diverse domains such as science, engineering, AI and machine learning, and health technology. The Web of Science database is a cross-disciplinary resource, encompassing research papers from a broad spectrum of disciplines, including science, technology, art, and social science. These databases collectively provide extensive coverage of research, supplying researchers with valuable and insightful information.

2.1 Approach of Searching

The five databases that were taken into consideration underwent a thorough bibliographic search for academic papers written in English (SD, Scopus, IEEE, ACM Digital Library, and WoS). All scientific articles published between the beginning of scientific output and May 2022 were included in this search. Specifically, this search employed a Boolean query based on 'OR,' with 'declarative machine learning' and 'auto machine learning' as the chosen keywords.

2.2 Criteria of Insertion and Exclusion

The criteria used for article inclusion/selection are crucial aspects of this systematic literature review. The following criteria were employed for this research:

- Articles were required to be authored in English and presented in a journal or conference proceedings.
- Articles are needed to explore technologies such as declarative machine learning or auto machine learning.

Studies beyond the scope of this research were excluded based on the following criteria:

- Papers in languages other than English.
- Contributions that did not fall within the scope of declarative machine learning.

2.3 Study Selection

Consistent with previous publications [14, 15], this study conducted a systematic literature review using the preferred reporting items for systematic review statements. The method involved multiple processes, beginning with the elimination of duplicate papers. Titles and abstracts of the contributions were scanned using Mendeley software. This process included all authors, and many unrelated works were excluded. Any differences or disagreements among authors were resolved by the corresponding author. The third step involved a comprehensive review of the entire text, eliminating articles that did not encounter the previously stated requirements of inclusion (refer to Section 2.2). Two experts completed the filtering procedure to assess its effectiveness. Articles that satisfied the criteria were included in this study. The initial search yielded 201 results, with 49 from the Science Direct repository, 58 from a database of Scopus, 11 from Xplore of IEEE, 49 from the ACM digital library, and 34 from a repository of (WoS). The exploring encompassed all works issued between the commencement of research construction and September 28, 2023. The total number of articles was reduced to 53 after eliminating approximately 12 replicas from the five databases. Following a thorough and critical analysis of titles and abstracts, 148 research articles were found unqualified, and just 41 papers were considered appropriate and incorporated in the ultimate list of publications based on the norms of inclusion. The next section examines the challenges associated with machine learning and computing technology.

3. CHALLENGES OF ML

This section highlights the importance of bringing together different "traditional" ideas into a fresh approach called Declarative Learning Based Programming. Additionally, there is a need for in-depth investigations to understand the requirements and challenges of machine learning in terms of the languages, representations, and computational models, etc. that will support this new paradigm. This section systematically reviews the difficulties that hinder the emergence and development of Declarative ML or Auto ML technologies. It explores these challenges across several categories, including feature engineering and information extraction, the structure and activities of learning models, learning-based analysis and visualization, algorithms for analysis in data-based systems, the development of machine learning algorithms and systems, and declarative ML-based prediction.

3.1 Features engineering and information extraction

This category addresses the challenges of feature extraction methodologies in comparison to declarative feature definition methods. The ability to identify and extract features for learning models from various data sources is a primary goal of data interaction in learning-based applications. Extracting low-level characteristics from learning samples, such as a phrase's length or the lemma of its words, is a common issue of feature engineering. Other capabilities include choosing, prognostic, or linking features, from one extractor to another, for complicated and structured data, selecting features, and mapping features. This indicates that the two aforementioned problems of interacting with raw data, organizing it, and enquiring about the outcome structure are addressed by the method of feature extraction and engineering. While research has been conducted on each of them, a unifying framework and a programming environment that supports machine learning are still lacking [1].

In Verbeke et al. [16], the authors highlighted the challenges of contextual features for NLP problems, stating that, in addition to this relational learning technique, declarative feature definition allows for the inclusion of extra background knowledge, often necessary when addressing NLP challenges. A significant data management difficulty is controlling the feature selection procedure. Singh et al. [17] addressed the process of creating machine learning algorithms for applications in natural language processing (NLP), emphasizing its inherently iterative nature. It requires constant improvement in model selection, feature engineering, inference algorithm selection, hyper-parameter search, and error analysis. Current probabilistic programming languages (PPLs) only provide partial answers; most of them do not support widely used models like neural networks or matrix factorization, nor do they enable interactive and iterative programming, essential for the rapid construction of these models. Zhang et al. [18] focused on materialization optimizations based on managing feature selection. Choosing the best materialization approach is challenging for analysts as it depends on the feature selection task's reuse opportunities, the amount of error the analyst is willing to accept, and the characteristics of the data and computing node, including parallelism and data size. For this reason, the best materialization approach for an R script on one dataset might not be the best approach for the same task on a different dataset. Analysts find it challenging to select the ideal set of materialization optimizations as a result.

3.2 Learning model's structure and activities

This category focuses on the structure and activities of learning models, explaining challenges associated with automatic ML model selection activities, the gradual and dynamic nature of typical ML development, constraint problems, building complex AI systems, and task parallelism.

In terms of previous research on feature subset selection (FSS) and algorithm suggestion, a limitation is identified in the use of a singular learner for meta-modelling, constraining its abilities. Additionally, much meta-modelling in current literature relies on a single set of data characterization measures [19]. State et al. [20] highlighted the growing importance of explaining opaque machine learning (ML) models. Current AI (XAI) approaches, while in use, have drawbacks, including a lack of abstraction and user interaction, as well as inadequate integration of prior knowledge. Li et al. [21] discussed a multitenant scenario where resource allocation poses a crucial yet complicated issue, requiring a compromise between effectiveness and fairness. They formalize the problem of multi-tenant model selection to reduce overall user disappointment during automatic model selection activities. Xin et al. [22] addressed the acceleration of normal ML, emphasizing the current focus on accelerating workflows executed one time, neglecting the gradual and dynamic nature of typical ML development. Wang et al. [23] highlight the urgent need for declarative machine learning over-dispersed data platforms and argue that Datalog-based declarative abstractions are natural matches for machine learning, particularly in comparison to simpler applications already supported by BigDatalog. Spieker et al. [24] find challenges in supervised learning when approaching data-driven constraint solutions. They frame a constraint problem as a structured sequence-to-sequence (S2S) problem with bandit feedback, emphasizing the impracticality of deploying an RL-trained model due to potentially infeasible answers for real use cases. Kordjamshidi et al. [25] emphasize that building complicated AI systems requires significant effort in programming and expertise, addressing diverse learning and cognitive paradigms at a relatively low limit of declarative and abstraction. The lack of theoretical understanding or methods to abstract over these complexities necessitates substantial experimental exploration for selecting a model, selecting features, and parameters adjustment. Boehm et al. [26] discuss declarative ML system optimizations for data and task parallelism challenges, with

a focus on SystemML's emphasis on data parallelism. The main challenge lies in effectively integrating both forms of parallelism for various machine learning workloads and scripts.
To address these issues, considerable work has been given to automating or declaratively learning the procedures associated with the structure and activities of the Learning model. Automated machine learning, sometimes denoted as "Auto-ML," is the procedure that automates methods of the time-wasting and iterations involved in the development of machine learning models. The primary goal is to reduce user or individual efforts in developing precise prediction and estimation models, encourage early deployment of best solutions, and invest time and money while maintaining accuracy.

3.3 Learning-based analysis and visualization

This category emphasizes the challenges of ML languages related to visual analysis and information visualization. Li & and Ma [27] discussed the difficulty of integrating interactive visual analysis with machine learning techniques. Currently, available libraries of declarative programming and visualization toolkits do not support the machine learning techniques combination. According to Lekschas et al. [28], small multiples are microscopic visual information characteristics used in various fields. Dealing with a high number of small multiples complicates several analytical activities, such as inspection, comparison, navigation, and annotation.

3.4 Machine Learning Algorithms and Systems Development

This section delves into the challenges of developing machine learning algorithms and systems, covering various aspects such as the costs associated with a large class of ML algorithms, issues influencing the creation of declarative machine learning systems, difficulties in utilizing machine learning systems on the cloud, task parallelism challenges in ML systems, and the role of declarative machine learning languages in database systems to address knowledge base construction (KBC) problems.
Leskovec [29] emphasized the challenges of creating AI-powered solutions, highlighting that it takes highly experienced teams months, or even years, to train and deploy machine learning models in the real world. Streamlining the workflow for machine learning is essential to make AI more reachable to a broader range of users. Ghoting et al. [30] discussed the cost implications of executing a large class of ML algorithms as MapReduce jobs of low-level on diverse data and machine cluster volumes, emphasizing the potential prohibitive costs. Molino et al. [2] addressed issues influencing the creation of declarative machine learning systems, including challenges related to choices during system construction, the tendency for "New Model-itis," organizational gaps, a scarcity of expertise, process sluggishness, and a diverse set of stakeholders. Zhang et al.[31] argued that modern machine learning systems, despite recent advances, remain challenging for users without a computer science background. They explored the advantages and losses for consumers using declarative machine learning clouds compared to non-declarative systems. Boehm et al. [26] discussed challenges in data and task parallelism in machine learning systems built on MapReduce, highlighting performance issues and the lack of default support for task parallelism. They emphasized the importance of addressing efficiency and scalability across various job sizes. De Sa et al.[32] focused on using declarative machine learning languages in database systems to address knowledge base construction (KBC) problems. They highlighted the challenges of adding information from unstructured data sources to a relational database, known as dark data extraction or KBC. Consuegra-Ayala et al. [33] emphasized the need for developing AutoML methods, showcasing advancements in Automatic Machine Learning (AutoML) tools like Auto-Weka, Auto-sklearn, and Auto-Keras. These tools efficiently determine the optimal mix of algorithms and hyperparameters, reducing the time researchers spend on well-researched issues. Nunes de Oliveira et al. [34] revealed that most AutoML systems use Evolutionary Algorithms (EA) or Bayesian Optimization (BO) to identify optimal solutions, with challenges related to the feasibility of evaluating every potential pipeline in combinatorial optimization problems.
Hence, to address these difficulties, both AutoML and Declarative ML share the common goal of streamlining the machine learning process.

3.5 Declarative ML-based Prediction

This category outlines challenges related to predicting quantities of interest, monitoring and forecasting health problems in people, and addressing issues of machine learning language for distributed computing platforms in addition to other problems in machine learning studies.
Unluckily, a bunch of crucial quantities of interest (QoI) cannot be immediately measured using sensors, like the weight of an aeroplane, which could lead to accidents. Gurney et al. [35] addressed these challenges of predicting quantities of interest stating that utilizing cognitive models to estimate QoI from other aircraft sensor data exposes aeroplanes or objects of interest to risk. However, direct measurement of certain QoI, such as aircraft weight, can be challenging, potentially leading to accidents.

In another scenario, when using an algorithm to monitor and forecast health problems in people, gaps will eventually appear. This consideration encompasses entirely the stages in machine learning model development, with a particular emphasis on the initial stages, including identifying the problem and collecting the data and stages of preparation [36]. Gao et al.[37] emphasized the necessity for a declarative machine learning language for distributed computing platforms, outlining challenges faced in meeting such a requirement. Some problems include the difficulty of creating an ML or statistical application that extracts useful information from a large amount of data, particularly when transitioning to a distributed or parallel implementation. This becomes especially challenging when using dataflow platforms like Hadoop, Spark, DryadLinq, or Flink. Musigmann et al. [38] highlighted challenges associated with machine learning in medical research, despite consistent growth in ML research. Effectively using these techniques requires specialized expertise, involving time-consuming steps such as data division, feature preselection, multivariate feature selection, hyperparameter optimization, and model construction and selection. Issues like overfitting and underfitting further complicate the process. The potential application of machine learning frameworks on transcriptomic data for finding biomarker signatures in predicting binary classification (yes/no) of patient survival is believed to be promising. However, this method is not yet widely used to improve therapy prognostics. In bioinformatics, various machine-learning methods can group biomarkers and enhance prediction power, but the abundance of parameter permutations makes finding the best models challenging and time-consuming [39]. Despite advancements in performance prediction, most modeling-related activities in machine learning remain challenging for non-experts, including algorithm selection, model optimization, and additional tasks [40]. Thus, coping with the aforementioned challenges, AutoML and Declarative ML represent significant steps towards making machine learning more accessible to a broader range of users.

4. CONCLUSION

ML technologies have turned into a potent instrument for deriving conclusions and insights from data. However, creating and implementing machine learning models can be challenging and time-wasting, often necessitating programming and data science knowledge. This paper introduced a systematic review of the latest research to highlight the challenges of machine learning languages, algorithms, and systems, emphasizing the need to maintain some declarative ideas. This study contributed to identifying existing challenges in machine learning, aligning with the emerging concepts of AutoML and Declarative ML that can simplify ML technologies and make them more effective for end-users of AI. The future work of this study is to investigate particular issues in ML technologies related to AutoML and Declarative ML and their proposed solutions, highlighting still unresolved problems.

Funding

This investigation was carried out without the financial backing of any institutions or sponsors.

Conflicts of Interest

The authors affirm that there are no conflicts of interest connected to this study.

Acknowledgment

The authors extend heartfelt gratitude to the institution for their unwavering support and invaluable contributions throughout this research endeavor.

References

[1] P. Kordjamshidi, D. Roth, and K. Kersting, "Systems AI: A Declarative Learning Based Programming Perspective," in IJCAI, 2018, pp. 5464-5471.
[2] P. Molino and C. Ré, "Declarative machine learning systems," Communications of the ACM, vol. 65, no. 1, pp. 42-49, 2021.
[3] A. Alexandrov et al., "The stratosphere platform for big data analytics," The VLDB Journal, vol. 23, pp. 939-964, 2014.
[4] M. Zaharia et al., "Resilient distributed datasets: A {Fault-Tolerant} abstraction for {In-Memory} cluster computing," in 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12), 2012, pp. 15-28.
[5] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," Communications of the ACM, vol. 51, no. 1, pp. 107-113, 2008.
[6] M. Boehm, A. V. Evfimievski, N. Pansare, and B. Reinwald, "Declarative machine learning-a classification of basic properties and types," arXiv preprint arXiv:1605.05826, 2016.
[7] X. Meng et al., "Mllib: Machine learning in apache spark," The journal of machine learning research, vol. 17, no. 1, pp. 1235-1241, 2016.

[8] R. Anil et al., "Apache mahout: Machine learning on distributed dataflow systems," The Journal of Machine Learning Research, vol. 21, no. 1, pp. 4999-5004, 2020.

[9] J. Hellerstein et al., "The MADlib analytics library or MAD skills, the SQL," arXiv preprint arXiv:1208.4165, 2012.

[10] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton, "MAD skills: new analysis practices for big data," Proceedings of the VLDB Endowment, vol. 2, no. 2, pp. 1481-1492, 2009.

[11] P. Molino and C. Ré, "Declarative Machine Learning Systems: The future of machine learning will depend on it being in the hands of the rest of us," Queue, vol. 19, no. 3, pp. 46-76, 2021.

[12] D. J.-L. Lee and S. Macke, "A Human-in-the-loop Perspective on AutoML: Milestones and the Road Ahead," IEEE Data Engineering Bulletin, 2020.

[13] R. Elshawi and S. Sakr, "Automated machine learning: Techniques and frameworks," in Big Data Management and Analytics: 9th European Summer School, eBISS 2019, Berlin, Germany, June 30–July 5, 2019, Revised Selected Papers 9, 2020, pp. 40-69: Springer.

[14] C. Sohrabi et al., "PRISMA 2020 statement: What's new and the importance of reporting guidelines," vol. 88, ed: Elsevier, 2021, p. 105918.

[15] K. W. Khaw, A. Alnoor, H. Al-Abrrow, V. Tiberius, Y. Ganesan, and N. A. Atshan, "Reactions towards organizational change: a systematic literature review," Current Psychology, vol. 42, no. 22, pp. 19137-19160, 2023.

[16] M. Verbeke, P. Frasconi, K. De Grave, F. Costa, and L. De Raedt, "klognlp: Graph kernel–based relational learning of natural language," in Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2014, pp. 85-90.

[17] S. Singh, T. Rocktäschel, L. Hewitt, J. Naradowsky, and S. Riedel, "WOLFE: an nlp-friendly declarative machine learning stack," in Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, 2015, pp. 61-65.

[18] C. Zhang, A. Kumar, and C. Ré, "Materialization optimizations for feature selection workloads," ACM Transactions on Database Systems (TODS), vol. 41, no. 1, pp. 1-32, 2016.

[19] I. Khan, X. Zhang, R. K. Ayyasamy, and R. Ali, "AutoFe-Sel: A Meta-learning based methodology for Recommending Feature Subset Selection Algorithms," KSII Transactions on Internet & Information Systems, vol. 17, no. 7, 2023.

[20] L. State, S. Ruggieri, and F. Turini, "Declarative Reasoning on Explanations Using Constraint Logic Programming," in European Conference on Logics in Artificial Intelligence, 2023, pp. 132-141: Springer.

[21] T. Li, J. Zhong, J. Liu, W. Wu, and C. Zhang, "Ease. ml: Towards multi-tenant resource sharing for machine learning workloads," Proceedings of the VLDB Endowment, vol. 11, no. 5, pp. 607-620, 2018.

[22] D. Xin, L. Ma, J. Liu, S. Macke, S. Song, and A. Parameswaran, "Helix: Accelerating human-in-the-loop machine learning," arXiv preprint arXiv:1808.01095, 2018.

[23] J. Wang, J. Wu, M. Li, J. Gu, A. Das, and C. Zaniolo, "Formal semantics and high performance in declarative machine learning using Datalog," The VLDB Journal, vol. 30, no. 5, pp. 859-881, 2021.

[24] H. Spieker, "Towards sequence-to-sequence reinforcement learning for constraint solving with constraint-based local search," in Proceedings of the AAAI Conference on Artificial Intelligence, 2019, vol. 33, no. 01, pp. 10037-10038.

[25] P. Kordjamshidi, D. Roth, and K. Kersting, "Declarative learning-based programming as an interface to AI systems," Frontiers in artificial intelligence, vol. 5, p. 755361, 2022.

[26] M. Boehm et al., "Hybrid parallelization strategies for large-scale machine learning in systemml," Proceedings of the VLDB Endowment, vol. 7, no. 7, pp. 553-564, 2014.

[27] J. K. Li and K.-L. Ma, "P6: A declarative language for integrating machine learning in visual analytics," IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 2, pp. 380-389, 2020.

[28] F. Lekschas, X. Zhou, W. Chen, N. Gehlenborg, B. Bach, and H. Pfister, "A generic framework and library for exploration of small multiples through interactive piling," IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 2, pp. 358-368, 2020.

[29] J. Leskovec, "Databases as Graphs: Predictive Queries for Declarative Machine Learning," in Proceedings of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, 2023, pp. 1-1.

[30] A. Ghoting et al., "SystemML: Declarative machine learning on MapReduce," in 2011 IEEE 27th International conference on data engineering, 2011, pp. 231-242: IEEE.

[31] H. Zhang, L. Zeng, W. Wu, and C. Zhang, "How good are machine learning clouds for binary classification with good features? extended abstract," in Proceedings of the 2017 Symposium on Cloud Computing, 2017, pp. 649-649.

[32] C. De Sa et al., "Deepdive: Declarative knowledge base construction," ACM SIGMOD Record, vol. 45, no. 1, pp. 60-67, 2016.

[33] J. P. Consuegra-Ayala, Y. Gutiérrez, Y. Almeida-Cruz, and M. Palomar, "Intelligent ensembling of auto-ML system outputs for solving classification problems," Information Sciences, vol. 609, pp. 766-780, 2022.

[34] D. Nunes de Oliveira and L. H. d. C. Merschmann, "An Auto-ML Approach Applied to Text Classification," in Proceedings of the Brazilian Symposium on Multimedia and the Web, 2022, pp. 108-116.

[35] S. Gurny, J. Falvo, and C. Varela, "Aircraft Weight Estimation During Take-off Using Declarative Machine Learning," in 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), 2020, pp. 1-10: IEEE.

[36] D. Casacuberta, A. Guersenzvaig, and C. Moyano-Fernández, "Justificatory explanations in machine learning: for increased transparency through documenting how key concepts drive and underpin design and engineering decisions," Ai & Society, pp. 1-15, 2022.

[37] Z. J. Gao, S. Luo, L. L. Perez, and C. Jermaine, "The BUDS language for distributed bayesian machine learning," in Proceedings of the 2017 ACM International Conference on Management of Data, 2017, pp. 961-976.

[38] M. Musigmann et al., "Testing the applicability and performance of Auto ML for potential applications in diagnostic neuroradiology," Scientific reports, vol. 12, no. 1, p. 13648, 2022.

[39] R. J. Pais, F. Lopes, I. Parreira, M. Silva, M. Silva, and M. G. Moutinho, "Predicting Cancer Prognostics from Tumour Transcriptomics Using an Auto Machine Learning Approach," in Medical Sciences Forum, 2023, vol. 22, no. 1, p. 6: MDPI.

[40] C. Yang, S. An, B. Qiao, P. Guan, D. Huang, and W. Wu, "Exploring the influence of COVID-19 on the spread of hand, foot, and mouth disease with an automatic machine learning prediction model," Environmental Science and Pollution Research, vol. 30, no. 8, pp. 20369-20385, 2023.