How automated machine learning help business data science teams?

Ebadi, Ashkan; Gauthier, Yvan; Tremblay, Stéphane; Paul, Patrick

Publisher's version / Version de l'éditeur:
https://doi.org/10.1109/ICMLA.2019.00196
International Conference on Machine Learning and Applications (ICMLA 2019), pp. 1186-1191, 2019-12-19

How can automated machine learning help business data science teams?

Ashkan Ebadi
Digital Technologies Research Centre
The National Research Council Canada
Montréal, Canada
ashkan.ebadi@nrc-cnrc.gc.ca

Yvan Gauthier
DRDC Centre for Operational Research and Analysis
Department of National Defence
Ottawa, Canada
yvan.gauthier@forces.gc.ca

Stéphane Tremblay
Digital Technologies Research Centre
The National Research Council Canada
Ottawa, Canada
stephane.tremblay@nrc-cnrc.gc.ca

Patrick Paul
Digital Technologies Research Centre
The National Research Council Canada
Ottawa, Canada
patrick.paul@nrc-cnrc.gc.ca

***Abstract*—Artificial intelligence and machine learning have attracted the attention of many commercial and non-profit organizations aiming to leverage advanced analytics, in order to provide a better service to their customers, increase their revenues through creating new or improving their existing internal processes, and better exploit their data by discovering complex hidden patterns. Such advanced solutions require data scientists with rare (and generally expensive) skill sets. Moreover, such solutions are often perceived as complex black boxes to decision-makers. Automated machine learning tools aim to reduce the expertise gap between the technical teams and stakeholders involved in business data science projects, by reducing the amount of time and specialized skills required to generate predictive models. We systematically benchmarked five automated machine learning tools against seven supervised learning problems of a business nature. Our results suggest that such tools, in fully automated mode, must be used cautiously, only where predictive models support low-impact decisions and do not need to be explainable, and only by data scientists capable to ensure that all phases of the data mining process have been performed adequately.**

Keywords—automated machine learning, augmented analytics, decision making, decision support, data science, industry, artificial intelligence, supervised learning

I. Introduction

Machine learning (ML) and artificial intelligence (AI) have already progressed from theoretical academic research projects to enterprise applications. Commercial and non-profit organisations are now trying to incorporate AI to their data and analytics infrastructure in order to become more effective and gain business advantage. Despite high expectations for AI, business enterprises found it difficult to fully embrace it in their processes. Hence, the gap between ambition and implementation is significant in most organizations [1]. High data volume, variety, and complexity have made it difficult, if not impossible, to be efficiently processed by conventional tools [2].

However, "big" data present an opportunity for enterprises to transform their business leveraging advanced analytics [3]. In industry, data-driven decision making is now recognized as one of the drivers for improved profitability [4]. The growing attention towards data-as-an-asset on one hand, and the skills and resources required to gain business advantage from data, on the other hand, have had deep implications in how companies are setting strategies towards incorporating AI.

The term *data scientist* was coined by D. J. Patil in 2008, referring to data analytics experts who work on data applications with a massive and immediate organizational impact [5]. By this definition, data scientists translate business problems to data science projects and solve them by applying a scientific method to voluminous sets of data of different types [6]. Proficient data scientists have a broad set of skills including but not limited to domain knowledge, programming, statistical and analytical skills, communication expertise, and teamwork [3], [7]. Such requirements have made proficient data scientists rare commodities, and because of high demand, they are becoming more expensive [8].

Growing data, shortage of skillful data scientists, along with the market push towards intelligent solutions are some of the main factors that have made automated machine learning (AutoML) tools attractive to many businesses. Although there are several AutoML tools now available, they all claim to produce valuable results with the least effort required [9]. The main goal of AutoML tools is to reduce the amount of time and specialized skills required to generate, deploy and maintain predictive models, by automating the most repetitive steps of the data science lifecycle. This automation could help data scientists to accelerate the pace of these steps and focus more on other important aspects of analytics [10]. But due to the short supply of data scientists and the need for boosting productivity and efficiency, AutoML tools are often perceived in the industry as a shortcut to implementing AI capabilities or even an alternative for skilled data scientists.

In this study, we benchmarked five automated machine learning (ML) tools, in fully automated mode, and evaluated their model building performance against human-crafted data science solutions. This systematic analysis provides a comprehensive assessment of select AutoML solution alternatives and effort-performance trade-offs. Moreover, it helps to figure out to what extent businesses can or should rely on AutoML tools.

II. Datasets

We considered seven public datasets of business nature to perform the analysis. The target datasets were acquired from two main sources: 1) Kaggle, which is an online community of data science practitioners and a place to do data science projects/competitions, and 2) Scikit-Learn [11] datasets, that embed some small datasets, mostly used by the machine learning community for educational purposes. The datasets were selected based on multiple criteria, as defined below, aiming to cover a diverse set of issues in common data science projects:

- Skills required to explore and extract insight from the data, such as creative feature engineering, missing data handling, etc.
- Ability to apply supervised learning.
- Amount of data-related issues, e.g., outliers, missing data, data leakage.
- Existence of a benchmark optimal solution.
- Diversity of datasets in terms of the number of observations and features.

One may note that these public datasets are still relatively tidy and ready for analysis. In reality, business datasets are often much more complex, incomplete and not always aligned with analytical objectives. Table I summarizes the select datasets. All datasets, except for Boston, had a labelled training data along with an unlabelled test set. Boston data points were all labelled.

TABLE I. DATASETS.

Dataset	Data Dimension		Source	Problem Type
	#Obs.	*#Feat.*		
Springleaf [12]	290,463	1,933	Kaggle	Binary classification
Otto Group [13]	206,246	94	Kaggle	Multi-class classification
Santander Bank [14]	151,838	370	Kaggle	Binary classification
Rossmann [15]	1,058,297	8	Kaggle	Regression
Liberty Mutual [16]	101,999	33	Kaggle	Regression
Ames [17]	2,919	80	Kaggle	Regression
Boston [18]	506	14	Scikit-learn	Regression

III. AUTOML TOOLS BENCHMARKED

A. H2O AutoML (version 3.2.1.1)

H2O is an open-source, in-memory, distributed machine learning, deep learning, and predictive analytics platform that allows the user to build machine learning models on large-scale data [19]. H2O AutoML can be used for automating the machine learning workflow through automatically training (ensemble) models on the given dataset. In the fully automated mode, the user often only needs to provide the dataset and identify the target variable. There is also a possibility of specifying a time constraint or a limit on the number of total models trained. H2O massively uses computing and memory resources to build the model [9].

B. The auto_ml package (version 2.9.10)

The auto_ml package enables the user to automate the machine learning model building process. It could also perform feature engineering over data of different types such as date and text. Moreover, it provides hyperparameter optimization, feature selection, and feature scaling in the automated process. As part of the model fitting process, auto_ml needs the user to at least identify the target variable. Ideally, auto_ml requires the type of each feature as input to process it correctly [9].

C. TPOT (version 0.9.5)

TPOT was developed in the epistasis research lab at the University of Pennsylvania and is still under active development. It is a Python tool that optimizes machine learning pipelines using genetic programming. The main focus of TPOT is on the automation of feature preprocessing, feature selection, feature construction, hyper-parameter tuning, and model building. TPOT outputs the Python code of the best pipeline after the automatic search is over [20].

D. SAP Automated Analytics (version 3.3)

SAP Automated Analytics is the automated module of SAP Predictive Analytics Desktop, a business intelligence software from SAP, a German-based software corporation. The software aims to help organizations analyse large datasets by automating supervised learning. There is also an expert mode with a guided workflow allowing users to choose predictive functions according to the use case. SAP Predictive Analytics is a commercial product, offering a free 30-day trial.

E. Auto-sklearn (version 0.4.2)

Auto-sklearn wraps the scikit-learn framework to automatically create a machine learning pipeline. The package lacks the ability to process text input and it cannot automatically distinguish between numerical and categorical features. Therefore, it cannot be used in a fully automated mode as, for example, manual integer encoding of categorical features or identifying numerical features before the model building would be required. Auto-sklearn uses an optimization framework that implements a Bayesian search along with a racing mechanism to find the best model [21]. Therefore, the performance of the tool is highly dependent on the dataset characteristics, such as size and complexity, as well as the time limit for training models. Table II summarizes the examined AutoML tools.

TABLE II. AUTOML TOOLS.

Tool (Version)	Open-source	License	GitHub # Stars[a]
H2O AutoML (3.2.1.1)	Yes	Apache	4,348
Auto_ml (2.9.10)	Yes	MIT	1,338
TPOT (0.9.5)	Yes	GNU Lesser General Public	6,308
SAP Predictive Analytics software (3.3)	No	30-day free trial	NA
Auto-sklearn (0.4.2)	Yes	BSD-3 Clause	3,936

a. Until October 2019.

IV. METHODOLOGY

We tested the select AutoML tools in fully automated mode, i.e., taking the functions with default parameter values and applying them to the input data, against each other as well as human-crafted data science solutions made by one of our data scientists. We also benchmarked the results with the best possible human solution, indicated by Kaggle leaderboard scores.

A. Experiments

To be fair against the examined tools, we tested all models on exactly the same data splits. We performed two main experiments to test the models' performance as follows:

1. We trained the model on the entire training data and tested its performance on the unlabelled test set.[1] We

[1] We did this experiment for all datasets except for Boston dataset as it did not have unlabelled test dataset.

limited models to be built in three hours. The results were submitted to Kaggle to obtain the leaderboard score.

2. We only considered the labelled training datasets and randomly split them into 80%-20% training and validation sets. We did random split five times, hence, for each dataset, we created five different training and validation splits. The models were then built using the new training data, within a one-hour time limit, and were tested on the validation sets.

The experiment design allowed us to assess the models' performance using six different runs on each dataset, i.e., one run on the entire training data and five runs on the training data random splits. In total, we ran more than 350 experiments.

B. *Human-crafted models*

We followed a modified CRoss-Industry Standard Process for Data Mining (CRISP-DM) process [22] (CRISP-DM minus the deployment step) to build the human-crafted solutions (Fig. 1). The data scientist respected the same time limits as the AutoML tools, i.e., one and three hours, to build the models. But he spent up to fifteen hours on each dataset to understand the problem, explore the data, and build informative features.

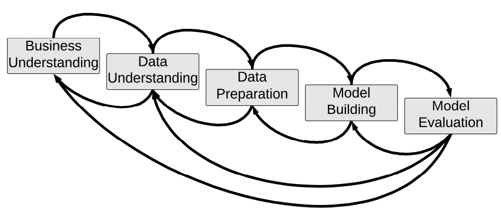

Fig. 1. The human-crafted data science pipeline.

C. *Performance metrics*

A performance metric was considered for each examined dataset based on the dataset characteristics and type of analytics. Table III lists the performance metrics for each dataset, based on the metrics used in Kaggle leaderboards.

TABLE III. PERFORMANCE METRICS.

Dataset	Performance Metric
Springleaf	Area under receiver operating characteristic curve (AUC)
Otto Group	Multi-class log loss
Santander Bank	AUC
Rossmann	Root mean square percentage error
Liberty Mutual	Normalized Gini index
Ames	Root mean square logarithmic error
Boston	Root mean square error

D. *Computing system configuration*

All the experiments were done on an HP EliteBook 820 G3 laptop with an Intel Core i5 CPU (2.4 GHz, 2 cores), 16 gigabytes RAM, and 500 gigabytes storage capacity.

V. RESULTS

A. *Dataset complexity vs. AutoML tools failures*

The examined datasets, as defined in Section II, were composed of four regression and three classification problems of diverse complexity. We defined a complexity score for the classification problems as below:

$$C_i^c = \sum \left(rank\left(\log\frac{mc_i}{r_i}\right) + rank\left(\log\frac{r_i}{f_i}\right) + rank\left(\log\frac{r_i}{tr_i}\right) + D_i + M_i\right) \quad (1)$$

In (1), C_i^c is the complexity score for classification dataset i, *rank* function returns the rank of each element in descending order, mc_i is the minimum number of observations per class in dataset i, r_i is the number of observations in the training dataset i, f_i is the number of features in the training dataset i, tr_i is the number of observations in the test dataset i. D_i is the dirtiness of dataset i and is defined as sum of four binary variables indicating whether the dataset has outliers, missing data, errors, and leakage. M_i is a binary variable indicating if dataset i needs intensive data manipulation and reframing. Higher C^c values indicate more complexity. For the regression dataset i, we used a similar complexity score, named C_i^r, as defined in (2).

$$C_i^r = \sum \left(rank\left(\log\frac{r_i}{f_i}\right) + rank\left(\log\frac{r_i}{tr_i}\right) + D_i + M_i\right) \quad (2)$$

We defined failure if an automated tool could not finish model building process within the time limit for any reason such as lack of time, memory, or limited tool functionality, and then counted the number of failures for each AutoML tool in the performed experiments. Fig. 2 shows total number of failures for the examined tools in classification and regression problems. The maximum number of failures per tool in classification problems was 18 (i.e., 6 experiments for each of the 3 datasets), and for regression problems was 23 (i.e. 6 experiments for 3 of the datasets plus 5 experiments for Boston dataset). For small, low-dimensional datasets such as Boston and Ames housing datasets, all AutoML tools generated predictions. Interestingly, TPOT failed to build the model within the time limit in all the classification experiments. The score meter in the figure shows the

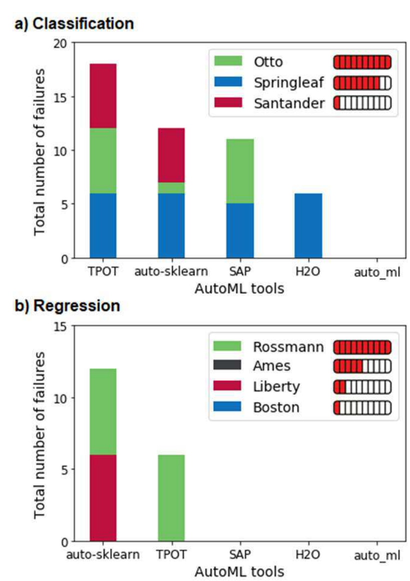

Fig. 2. Number of failures vs. dataset complexity: a) classification, b) regression datasets. The meter in front of the dataset names in the legend shows the complexity score calculated for the given dataset.

complexity of the given dataset, as defined in (1) and (2). It is also observed that, in general, the number of failures was higher in classification problems.

We further investigated the relation between AutoML tools' total number of failures per dataset with dataset complexity and dimension, defined as the number of rows times number of features, scaled in [1, 10] range. The relation between dataset complexity score and AutoML tools total number of failures was not very strong, however as seen in Fig. 3, number of failures is significantly and positively correlated with dataset dimension.

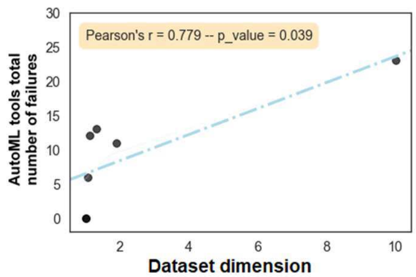

Fig. 3. Total number of automated ML tools failures per dataset versus dataset dimension scaled in [1, 10] range. The dashed line represents the regression line.

B. Comparing AutoML tools performance

We evaluated the performance of the AutoML tools on the examined regression and classification datasets. The performance was evaluated using the metrics defined in Table III. For each dataset we built models over six runs, except for the Boston dataset for which we did five.[2] We then calculated relative performance scores such that the AutoML tool with the highest score in each run obtained score of 100 and the others' scores were re-calculated relative to the top score. The final score was calculated by averaging the scores over the experiments. For example, if the final score for a tool is 100, it means that the tool outperformed all others in all runs. And, if the top final score for a dataset is lower than 100, e.g., in Boston and Ames, it indicates that none of the tools outperformed all the other tools in all runs. As seen in Fig. 4, the predictive models generated by H2O outperformed those of other tools for six out of seven problems. However, auto_ml was the only tool that was able to generate predictions for all the problems, generally faster than other tools, but generally less accurate than H2O models.

C. Benchmarking AutoML tools against human

We also benchmarked the performance of AutoML tools against predictive models built by humans. For this purpose, the performance of the best AutoML tool in each dataset was compared to: 1) the performance of our team's models built within the defined time limit, and 2) the top leading scores in the respective Kaggle competitions.[3] The latter could be considered as a good indicator of the best results achievable through human-crafted machine learning as hundreds of teams typically enter Kaggle competitions and the winning models are highly-tuned and specifically designed to maximize the predictive performance. As seen in Fig. 5, human-crafted models outperformed those from the examined AutoML tools. Kaggle leading models performed the best overall the datasets while our data scientist's models came second.[4] AutoML models were all in the form of ensembles of ensemble models. The data scientist tried to build a simpler but more interpretable model within the given time limit.

As expected, the performance gap between human-generated and machine-generated models is dependent on the dataset characteristics. Although the examined datasets were relatively cleaner and better structured than those normally encountered in real-life ML projects, as discussed in Sections II and V, we intentionally chose them to cover different levels of data complexity, and to assess if AutoML tools performance is consistent across different dataset properties.

We further investigated the performance gap between humans and AutoML and how it relates to dataset complexity. Complexity was defined by (1) and (2) and re-scaled to [1, 10] range. The relation between the dataset dimension and performance gap was weak. But according to Fig. 6, a moderate positive relation was observed between the dataset complexity score and human-AutoML tool performance ratio. That is the performance of automated tools decreases as datasets become more complex.

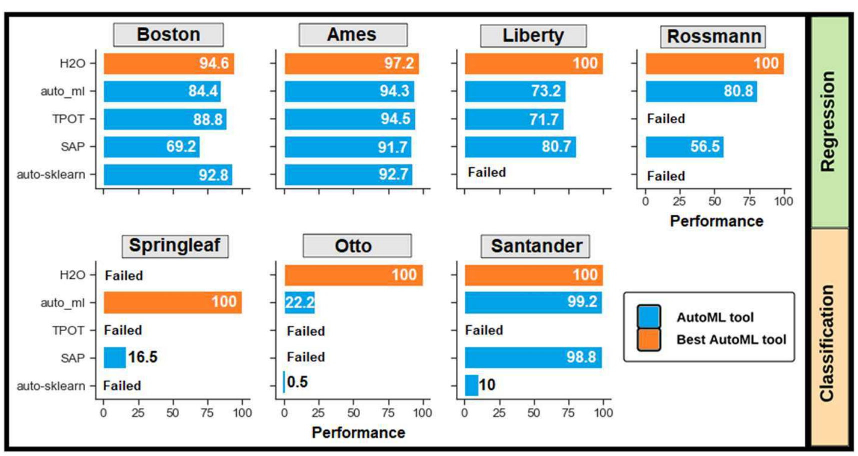

Fig. 4. Performance of automated machine learning tools.

[2] Please refer to the Section IV for details.

[3] To be able to compare with the Kaggle's leading scores, we trained the models on the entire training set and tested their performance on the test set.

[4] There was no unlabelled test set and hence no Kaggle scoreboard for Boston housing dataset.

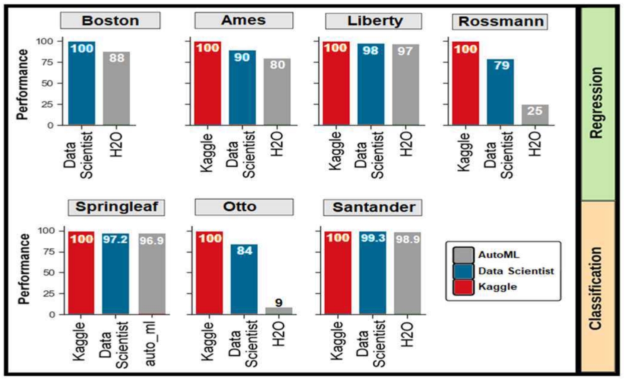

Fig. 5. Performance of the best automated machine learning tool versus human-crafted models.

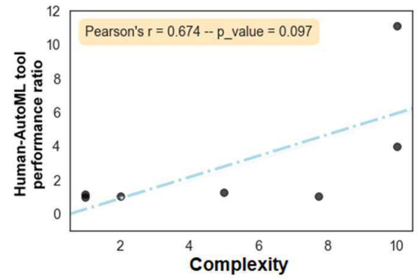

Fig. 6. Human-AutoML tool performance ratio versus dataset complexity score scaled in [1, 10] range. The dashed line represents the regression line.

VI. Discussion

Machine learning and artificial intelligence solutions have passed the research and development stage and proved to be successful in various applications. The drastic data growth rate on one hand, and organizations' eagerness to maximize benefits from their data, on the other hand, have challenged data science processes to be more productive, better connected to stakeholders' requirements, and easier to apply. Such improvements to processes may help data science teams to extract knowledge from data more efficiently and more effectively [10]. As one of many data science tools and packages, AutoML and other "augmented analytics" tools seek accelerating data science processes through automating some steps such as model building. In this paper, we focused on five AutoML tools, i.e., H2O AutoML, auto_ml, TPOT, auto-sklearn, and SAP Predictive Analytics software, and systematically benchmarked their performance (in fully automated mode) not only against each other but also against human-crafted models.

Machine learning solutions are often considered as complicated systems. The process of designing high-performing ML solutions is a mixture of art and science. Automated tools can definitely assist data scientists through facilitating different levels of system support for various steps in data science projects, especially, model selection. However, they may introduce additional cost and/or system complexity. In this study, we focused only on the model building capabilities of augmented analytics tools. The main challenge of business data science teams is often finding the right data, collecting it from multiple sources, and integrating the collected data.

In terms of modelling performance, we showed that AutoML tools still struggle competing with often simpler human-crafted models, even when processing low-complexity, clean datasets that are ready for modelling. Additionally, AutoML tools' risk of failure in the model building increases with the increase of dataset dimension. Of course, some measures could be taken to reduce such risk, e.g., relaxing the time limit or tuning AutoML tools parameters, but being aware of this issue is of high importance especially for business-related projects where datasets are often more complex and of higher dimension. That is, in real-life projects it is very likely that AutoML tools could not currently be used in fully automated mode as a self-service stand-alone tool.

Additionally, a positive relation was seen between dataset complexity and the human-AutoML performance ratio. Therefore, for complex datasets/problems such tools could not be employed without proper human supervision, following the data science project processes such as business understanding, data understanding, data exploration, and preparation. Although technology is the enabler of business processes automation, pure automation without a deep understanding of the problem is also against the human experience as we tend to automate things/processes that are repetitive, and we have good knowledge about them [23].

In data science competitions, achieving the highest performance is generally the objective. But in real-world situations, the performance-interpretability trade-off needs to be considered thoroughly. Predictive models generated by AutoML tools tend to be large stacks of ensemble models. Although these may perform well on some datasets, they produce results that are not easily explainable. As such, they are black boxes that may be useful to inform low-impact decisions (e.g., predicting when a piece of equipment should be sent for maintenance), but they should not be used to inform or automate any high-impact decision (e.g., decisions related to an organisation's employees).

The business has different drivers than the research community in incorporating AI capabilities, e.g., better and faster data-driven decision making, optimizing use of the collected data, etc. But most of these drivers link with maximizing revenue and profit which in a free market could involve highly time-sensitive processes or decisions to be made. Therefore, the focus in corporations is more on automating data science processes by using fully automated and/or drag-and-drop tools. That is, due to the high demand for analytics and at the same time lack of proficient data scientists, business is more pushing on automated ML tools without proper understanding of its capabilities and limitations. This could become more problematic if the company is not mature enough in terms of data science, software product development, and infrastructure requirements.

Automating the machine learning process blindly to reduce the cost of implementing it, as widely advertised in the business community (see [24], as an example), could be a risky strategy. For example, consider a company that is using an automated tool with a graphical interface to load, transform, and cleanse the data. If the tool is used in the automated mode and by a user without proper analytics knowledge, all missing values could be filled by a default value (e.g., zero), leading to possible information loss along the pipeline. Another example would be highly overfitted models with superb performance on the in-hand data, but poor on the unseen future data. This all calls for the necessity of analytics knowledge in performing data science projects.

Businesses may follow various analytics strategies based on their requirements and objectives (e.g., having a center of excellence for analytics, data science as a discipline) to handle large-scale projects. In such projects, the data scientist role often follows the definition proposed by [7], i.e., someone who has a diverse set of expertise including but not limited to computing theories, (advanced) algorithms, software engineering and system design concepts, data manipulation and management, familiarity with technology trends, data interpretation and story-telling skills, and even personal and social proficiency. Additionally, such a person needs to be able to assure performance, efficiency, scalability, and reliability of the proposed solution [7]. Such a complex profile is difficult to automate.

VII. Conclusion

Despite the fact that AutoML tools cannot replace human input (at least in the current state), AutoML tools remain useful aids to explore different modelling approaches, and to accelerate the modelling and deployment phases of the data mining process. Through automating repetitive and simple(r) tasks, data science teams could better concentrate on more complex tasks. Additionally, such tools could even assist data scientists to accelerate the model building process through building baseline models faster, and evaluating various analytics approaches/designs more quickly. Of course, AutoML tools' utility is expected to increase as their capabilities continue to progress. At the moment, AutoML tools must be used cautiously, only where predictive models support low-impact decisions and do not need to be explainable, and only by data scientists capable to ensure that all phases of the data mining process (not just modelling) have been performed adequately.

VIII. Limitations and Future Work

We used public competition/educational datasets to perform the analysis. Rules and assumptions in competitions might differ from the ones in real-world data science projects, as the goal in competitions is to maximize predictive performance. This is not always the preferred approach in business data science projects as several other factors, such as model interpretability, also play a role. Moreover, although the examined datasets cover a diverse set of data characteristics and complexity, they are, in general, much cleaner and better structured than the ones that could be found in real-world projects. Meanwhile, some of the examined AutoML tools were intensively used by several competition teams on the examined datasets. This may increase the risk of performance overestimation if the AutoML tool keeps a record of model building over datasets. Therefore, a possible future direction would be to do the same assessment on private datasets. This may provide a more concrete assessment of automated tools as the datasets would be completely new to the AutoML tools. Due to the project scope and duration, we put a time limit (one and three hours) on automated tools as well as hand-crafted ones to build the model. Further work could include only the high-performing automated tools and relax the time limit. We mainly assessed automated tools model building capabilities. Such tools could be also assessed against other steps in the data science pipeline, such as data collection, manipulation, and feature engineering.

References

[1] S. Ransbotham *et al*, "Reshaping business with artificial intelligence: Closing the gap between ambition and action," *MIT Sloan Management Review,* vol. 59, *(1),* 2017.

[2] S. Kaisler *et al*, "Big data: Issues and challenges moving forward," in *The 46th Hawaii International Conference on System Sciences,* 2013.

[3] M. A. Waller and S. E. Fawcett, "Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management," *Journal of Business Logistics,* vol. 34, *(2),* pp. 77-84, 2013.

[4] A. McAfee and E. Brynjolfsson, "Big data: the management revolution," *Harv. Bus. Rev.,* vol. 90, *(10),* pp. 60-6, 68, 128, 2012.

[5] D. Patil, *Building Data Science Teams.* 2011.

[6] T. H. Davenport and D. Patil, "Data scientist," *Harv. Bus. Rev.,* vol. 90, *(5),* pp. 70-76, 2012.

[7] C. Costa and M. Y. Santos, "The data scientist profile and its representativeness in the European e-Competence framework and the skills framework for the information age," *Int. J. Inf. Manage.,* vol. 37, *(6),* pp. 726-734, 2017.

[8] (Jan 17, 2019). ***Demand and Salaries for Data Scientists Continue to Climb***. Available: https://spectrum.ieee.org/view-from-the-valley/at-work/tech-careers/demand-and-salaries-for-data-scientists-continue-to-climb.

[9] A. Balaji and A. Allen, "Benchmarking Automatic Machine Learning Frameworks," *arXiv Preprint arXiv:1808.06492,* 2018.

[10] G. Ferruzza, M. Madireddy and A. Sharma, "Automated machine learning: An agile approach to AI deployment," Targetbase, 2018.

[11] F. Pedregosa *et al*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research,* vol. 12, *(Oct),* pp. 2825-2830, 2011.

[12] Springleaf (2015). ***Springleaf Marketing Response*** [Kaggle Datasets - Springleaf]. Available: https://www.kaggle.com/c/springleaf-marketing-response/data.

[13] Otto Group (2015). ***Otto Group Product Classification Challenge*** [Kaggle Datasets - Otto Group]. Available: https://www.kaggle.com/c/otto-group-product-classification-challenge/data.

[14] Santander Bank (2016). ***Santander Customer Satisfaction*** [Kaggle Datasets - Santander Bank]. Available: https://www.kaggle.com/c/santander-customer-satisfaction/data.

[15] Rossmann (2015). ***Rossmann Store Sales*** [Kaggle Datasets - Rossmann Store]. Available: https://www.kaggle.com/c/rossmann-store-sales/data.

[16] Liberty Mutual (2015). ***Liberty Mutual Group: Property Inspection Prediction*** [Kaggle Datasets - Liberty Mutual]. Available: https://www.kaggle.com/c/liberty-mutual-group-property-inspection-prediction/data.

[17] Ames (2011). ***House Prices: Advanced Regression Techniques*** [Kaggle Datasets - Ames Housing]. Available: https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data.

[18] Boston (1978). ***Boston Housing Dataset*** [Scikit-Learn Datasets - Boston Housing]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html.

[19] S. Aiello *et al*, *Machine Learning with R and H2O,* 2016.

[20] R. S. Olson *et al*, "Automating biomedical data science through tree-based pipeline optimization," in *European Conference on the Applications of Evolutionary Computation,* 2016.

[21] M. Feurer *et al*, "Efficient and robust automated machine learning," in *Advances in Neural Information Processing Systems,* 2015.

[22] C. Shearer, "The CRISP-DM model: the new blueprint for data mining," *Journal of Data Warehousing,* vol. 5, *(4),* pp. 13-22, 2000.

[23] McKinsey (Jul 2016). ***Where machines could replace humans—and where they can't (yet)***. Available: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet.

[24] Forbes (Apr 17, 2018). ***Seven Affordable Ways To Incorporate Machine Learning And AI Into Your Business***. Available: https://www.forbes.com/sites/forbestechcouncil/2018/04/17/seven-affordable-ways-to-incorporate-machine-learning-and-ai-into-your-business/#533500b54baf.