AutoAI: Automating the End-to-End AI Lifecycle with Humans-in-the-Loop

Dakuo Wang†, Parikshit Ram†, Daniel Karl I. Weidele, Sijia Liu, Michael Muller, Justin D. Weisz, Abel Valente, Arunima Chaudhary, Dustin Torres, Horst Samulowitz, Lisa Amini
IBM Research - dakuo.wang@ibm.com † Equal contribution

ABSTRACT

Automated Artificial Intelligence and Machine Learning (AutoAI / AutoML) can now automate every step of the end-to-end AI Lifecycle, from data cleaning, to algorithm selection, and to model deployment and monitoring in the machine learning workflow. AutoAI technologies, initially aimed to save data scientists from the low level coding tasks, also has great potential to serve non-technical users such as domain experts and business users to build and deploy machine learning models. Researchers coined it as "democratizing AI", where non-technical users are empowered by AutoAI technologies to create and adopt AI models. To realize such promise, AutoAI needs to translate and incorporate the real-world business logic and requirements into the automation. In this Demo, we present a first of its kinds experimental system, IBM AutoAI Playground, that enables non-technical users to define and customize their business goals (e.g., Prediction Time) as constraints. AutoAI then builds models to satisfy those constraints while optimizing for the model performance (e.g., ROC AUC score). This Demo also showcases AutoAIViz, a Conditional Parallel Coordinates visualization feature, and a TrustedAI feature from two accepted IUI'20 papers.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization**; • **Computing methodologies** → **Continuous space search**; **Machine learning approaches**.

KEYWORDS

AutoAI, AutoML, parallel coordinates, human-AI collaboration, democratizing AI, business constraints, stakeholder constraints

ACM Reference Format:
Dakuo Wang, Parikshit Ram, Daniel Weidele, Sijia Liu, Michael Muller, Justin D. Weisz, Abel Valente, Arunima Chaudhary, Dustin Torres, Horst Samulowitz, Lisa Amini. 2020. AutoAI: Automating the End-to-End AI Lifecycle with Humans-in-the-Loop. In *25th International Conference on Intelligent User Interfaces Companion (IUI '20 Companion), March 17–20, 2020, Cagliari, Italy*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3379336.3381474

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
IUI '20 Companion, March 17–20, 2020, Cagliari, Italy
© 2020 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-7513-9/20/03.
https://doi.org/10.1145/3379336.3381474

1 INTRODUCTION AND BACKGROUND

More and more business and organizations are applying machine learning techniques to gain insights from data and to support decision making, and this is referred to "Data Science" [3]. This end-to-end AI lifecycle is labor intensive and requires diverse skill sets [5, 6]; thus, recently, Automated AI and Machine Learning techniques have been developed to automate various steps of the process, such as preparing data, selecting models, engineering features, or optimizing hyperparameters [2, 4]. In this paper, we refer these systems as **AutoAI** or AutoML.

Research suggest that AutoAI can speed up the AI-building process for data scientists users, and data scientists praise the potential future of work where human data scientists and AutoAI can hand over tasks to each other in the AI lifecycle [7]. But we know very little about how non-technical users (e.g., business owners and domain experts) may interact with AutoAI in the AI Lifecycle, despite these non-technical stakeholders are essential to the work too [11].

To enable non-technical users to use AutoAI, there are two major challenges: **Business Objectives**, and **Transparency**. Business objectives challenge refer to the tension that what non-technical users want (e.g., business KPIs and policy compliance) often are not aligned with what technical users want (e.g., model accuracy and training time) [5]. Transparency challenge refers to that non-technical users do not necessarily understand *how* and *why* the AutoAI "blackbox" generates certain results [8].

In this Demo, we showcase an experimental AutoAI system, IBM AutoAI Playground, that aims to empower the non-technical users to communicate their business objectives as constraints in the automation optimization formula. The Demo leverages on a state-of-the-art joint optimization algorithm (i.e., ADMM) to populate AI models that can maximize the model performance (i.e., data science objective) while also satisfy the business constraints (i.e., **business objectives**) [4]. In addition, the Demo integrates AutoAIViz [10] – to tackle the **transparency** and trust challenge [1]. This visualization, based on Conditional Parallel Coordinates [9], provides interactive views of categorical and continuous value dimensions, allowing the user to expose increasing amounts of detail added to a conventional Parallel Coordinates display.

2 IBM AUTOAI PLAYGROUND

In this Demo, we will showcase new AI Automation capabilities for enterprise users, including the ability to automatically generate optimized machine learning pipelines that take into account business and stakeholder constraints, such as prediction resources, prediction latency, or fairness metrics.

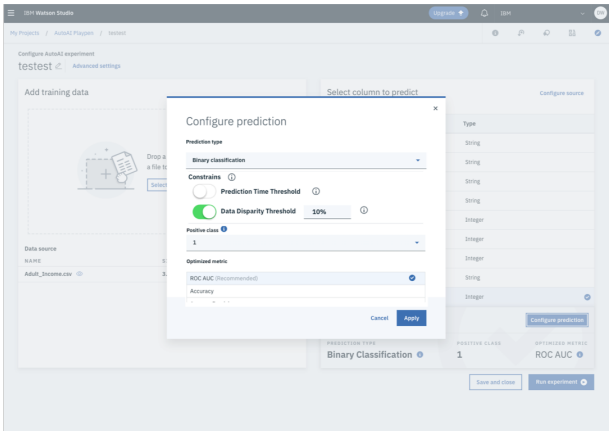


Figure 1: Screenshot showing AutoAI configuration UI. In this step, a user can upload a data set, specify the target column for training a model, and specify the optimization objective and business constraints by checking the toggle. As shown in this figure, the Data Disparity fairness constraint is enabled with a threshold of less than 10% and the Prediction Time constraint is disabled.

The entire user interaction with IBM AutoAI are simplified into two steps: a configuration step Figure 1 and a result step Figure 2. The users upload their dataset, define a problem type (i.e., binary classification, multi-class, or regression), then AutoAI can finish the rest of the work. Typically, during the process pipelines are being optimized towards a performance metric, in our case ROC AUC score is the default metric. In addition, users can also bound the AutoAI search space by setting constraints, such as *group disparity* against model bias, or *prediction runtime* for inference speed [4]. This extra capability is essential for non-technical users to leverage on AutoAI to solve their data science problems.

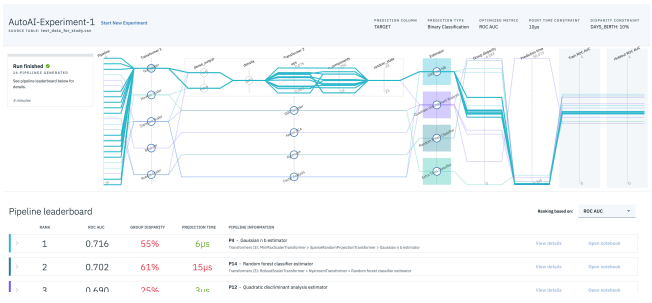


Figure 2: Screenshot showing AutoAI results UI, with AutoAIViz (top) and the leaderboard (bottom, partially shown). Each colored line running from left to right represents a machine learning pipeline, and corresponds to a row in the leaderboard. In this example, all pipelines consist of three steps: transformation step 1, transformation step 2, and estimator. Generated models are evaluated on four metrics: group disparity, prediction time, and ROC AUC scores on the training data, and on the holdout data.

Behind the scene, AutoAI first needs to decide on the sequence of the model pipeline generation process. AutoAI can choose a sequence of transformation steps before training an estimator, or it can simply decide not to use any transformation to generate new features. After the sequence is decided, AutoAI can search and decide on which particular transformer to be used inside each of the transformation step, and which estimator to be used inside the modeling step. Once that is done, AutoAI starts to fine-tune the hyperparameters of those chosen estimators and transformers.

Once AutoAI starts to generate results, the result page will instantly populate resulting model pipelines in progress(e.g., Figure 2). In AutoAIViz, each column represents one of the following attributes: *Pipeline ID, Transformer 1, Transformer 2, Estimator, Group Disparity, Prediction Runtime*, and *ROC AUC* on training and holdout split respectively. The next level of detail, for categorical choices, contains conditional axis per hyperparameter. The figure shows the visualization after the user expanded the Sparse Random Projection Transformer in the Transformer 2 step. Here, the visualization further allows to differentiate pipelines across the 5 hyperparameters *dense_output, density, eps, n_components* and *random_state*. This work is also accepted as an IUI'20 paper [10].

Towards the bottom of the result page Figure 2 we feature a pipeline leaderboard, showing pipeline configuration data sorted according to ROC AUC descendingly. Color indicates the choice of the estimator. For each pipeline row, from left to right, we display the rank, ROC AUC, Group Disparity, Prediction Time, ID and information on the steps.

REFERENCES

[1] Jaimie Drozdal, Justin Weisz, Dakuo Wang, Dass Gaurave, Bingsheng Yao, Changruo Zhao, Michael Muller, Lin Ju, and Hui Su. 2020. Exploring Information Needs for Establishing Trust in Automated Data Science Systems. In *IUI'20*. ACM, in press.

[2] Udayan Khurana, Deepak Turaga, Horst Samulowitz, and Srinivasan Parthasrathy. 2016. Cognito: Automated feature engineering for supervised learning. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 1304–1307.

[3] Sean Kross and Philip J Guo. 2019. Practitioners Teaching Data Science in Industry and Academia: Expectations, Workflows, and Challenges. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 263.

[4] Sijia Liu, Parikshit Ram, Deepak Vijaykeerthy, Djallel Bouneffouf, Gregory Bramble, Horst Samulowitz, Dakuo Wang, Andrew Conn, and Alexander Gray. 2019. An ADMM Based Framework for AutoML Pipeline Configuration. arXiv:cs.LG/1905.00424

[5] Yaoli Mao, Dakuo Wang, Michael Muller, Kush R. Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilovic. 2019. How Data ScientistsWork Together With Domain Experts in Scientific Collaborations. *Proceedings of the ACM on Human-Computer Interaction* 3, GROUP (Dec 2019).

[6] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *CHI'19*. ACM.

[7] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI. *CSCW'19* (2019).

[8] Qianwen Wang, Yao Ming, Zhihua Jin, Qiaomu Shen, Dongyu Liu, Micah J Smith, Kalyan Veeramachaneni, and Huamin Qu. 2019. Atmseer: Increasing transparency and controllability in automated machine learning. In *CHI'19*. ACM, 681.

[9] Daniel Karl I Weidele. 2019. Conditional Parallel Coordinates. In *2019 IEEE Visualization Conference (VIS)*. IEEE, 221–225.

[10] Daniel Karl I. Weidele, Justin Weisz, Erick Oduor, Michael Muller, Josh Andres, Alexander Gray, and Dakuo Wang. 2020. AutoAIViz: Opening the Blackbox of Automated Artificial Intelligence with Conditional Parallel Coordinates. In *IUI'20*. ACM, in press.

[11] Amy Zhang, Michael Muller, and Dakuo Wang. 2020. How do Data Science Workers Collaborate?: Roles,Workflows, and Tools. In *CSCW'20*. ACM, in press.