AutoML User Survey

, University of California, Berkeley, USA
, University of California, Berkeley, USA
, University of California, Berkeley, USA

CCS Concepts: • **Human-centered computing → Computer supported cooperative work**.

Additional Key Words and Phrases: Algorithmic Persona; YouTube; Metaphor; Folk Theories; Content Creators
, University of California, Berkeley, USA

ACM Reference Format:
. 2019. AutoML User Survey. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 219 (November 2019), ?? pages. https://doi.org/10.1145/3359321

1 INTRODUCTION

Machine learning development often involves an iterative, trial-and-error process of selecting the appropriate set of models, parameters, and settings that is optimal for a target application. This is a time-consuming and cubersome process that require users to experiment with a large set of modelling decisions and empirically evaluate the performance of the resulting model. To address these issues, in the past several years, we have seen a growing number of *AutoML systems* that aims to support and automate various aspects of the machine learning process. AutoML systems automatically search through a large space of possible modeling decisions, while optimizing for a set of desired objectives. AutoML systems leverages techniques from the field of meta-learning (learning-to-learn), such as bayesian optimization and neural architecture search, that enable efficient tuning and selection across a large number of model and parameters. The goals of these system is to abstract away the manual search and experimentation process to the machines so that users can focus on other parts of the machine learning workflow, such as engaging with business stakeholders or validating the robustness of the model before deployment. In this study, we aim to examine where Auto-ml fits within a typical machine learning workflow, the perception of auto-ml from users of auto-ml, and auto-ml practitioners' usage behavior.

We contribute to this research space by engaging with users of Auto-ml to inform the design of systems for individuals who can most benefit from Auto-ml. In addition, we fill in the gap in the understanding of how auto-ml can be designed for data scientists

Authors' addresses: eva.wu@berkeley.edu, University of California, Berkeley, Berkeley, California, USA;, University of California, Berkeley, Berkeley, USA, epedersen@berkeley.edu; University of California, Berkeley, Berkeley, USA, nsalehi@berkeley.edu.

2573-0142/2019/11-ART219 $15.00
https://doi.org/10.1145/3359321

to trust it enough to use it in practice by providing empirical evidence of perception of explainability, transparency and trust in auto-ml.

This is the first in-depth study of auto-ml practitioner to uncover their latent needs and wants. We reject the assumption that auto-ml design is to automate human workers, but rather to enhance humans' work. Here we extend prior research in calling for collaboration of human and AI and for AI to take on more an "augmentor" role rather than an "automator" role. The insights from the study will help guide the design of human-centered auto-ml systems.

2 RELATED WORK

Our work builds on top of three major areas of prior work: AutoML systems, human-centric view of AutoML and user studies of data science practitioners.

2.1 Automated Machine Learning (AutoML) Systems

AutoML technology is fast-growing and still in very much in its nascency, with a diverse set of commercial and academic offerings. Interfaces of AutoML tools ranges from GUI-based to programmatic, catering to different groups of users, including machine learning engineers, data scientists, and non-programming business users. AutoML tools vary in the degree of input customizability they offer, as well as their output model transparency. For example, some systems allow users to specify a range of models to select from (CITE) or metrics to optimize (CITE), while others (CITE) only allow users to specify the problem objective, such as classification or regression. In terms of the model output, some tools (CITE) exposes the model in its entirety to the user so that they can be fine-tuned further, while others (CITE google automl) simply return a monolithic black-box model that serves predictions. While AutoML tools have traditionally focused on automating the modelling phase of in the machine learning lifecycle, there has been a growing suite of AutoML tools that aims to democratize the end-to-end machine learning process, from data pre-processing to model building to post-processing. However, despite the excitement around this emerging technology, there have been little evaluation on how such tools are used in practice. In this paper, we sought to understand the adoption of AutoML tools, their usage, and the current bottlenecks that users are facing with these tools.

2.2 Human-Centric AutoML

Due to the lack of control and transparency in AutoML tools, recent research have investigated the human-centric aspect of AutoML [? ? ?]. These existing research explored data scientists' perceptions of transparency, trust and interpretability in AutoML systems (CITE) and designed novel visualization interfaces to improve transparency and control in AutoML [? ? ?]. These studies advocate for designing systems that facilitates constructive collaboration between humans and the AutoML system. Our work most closely resemble the work by the semistructured interviews on a AutoML prototype by Wang et al. [?], which revealed user's multifaceted perception of the role of AutoML as a collaborator or competitor. Our paper builds on this line of work by studying AutoML practitioners to understand how these tools are actually used in practice.

2.3 Empirical Studies of Data Practitioners

Empirical studies of data science and machine learning practitioner inform the design of data science tool and system. These research studies the work practices of specific groups of practitioners including software engineers using ML [?], non-experts ML users [?], and industry data scientists [?](dlee: I'm thinking maybe we shouldn't include the data science studies? do they talk about ML?). These existing studies focus on understanding how users develop ML models manually in their existing workflows and the challenge and bottlenecks they face in the machine learning process. Our study is the first that study practitioners who have experience using AutoML in real-world applications and contribute to an unprecedented understanding of work practices across all stages of the machine learning process through AutoML.

3 STUDY DESIGN

Our paper seeks to understand how auto-ml users incorporate auto-ml tools in their existing machine learning workflow and their perceptions of the tools.

3.1 Recruitment

We chose to focus on users who have used auto-ml for real-world use cases, in a diverse set of application domains. The participants either indicated that they had experience using at least one of the tools that we listed in the recruitment survey or they have used other tools that our team verified to be auto-ml [TODO: what is our criteria in defining auto-ml? - that the tool at least does automated model selection and hyperparameter tuning?].

We found participants in three ways: First, our personal connections (x participants). Second, we searched mentions of specific automated machine learning tools on twitter and invited the twitter users to fill out the recruitment form which asks whether they have experience using automated machine learning and sign up for an interview slot if the answer is affirmative (x participants). Third, we posted a recruitment message various relevant mailing lists and slack channels both in our university and in the industry.

3.2 Participants

We interviewed a total of 20 people (x male, x female; 3 White/Caucasian, 3 Asian, 2 South Asian, 1 Hispanic; aged 18 to 30, M = 21).

Our interview participants had an average of x years experience in machine learning and x years of experience in programming.

[TODO: add more quantitative measures this weekend, i.e. job titles, business sectors]

3.3 Interview Procedure

We began our study by conducting semi-structured interviews with users of automated machine learning for real-life use cases. This enabled us to establish a basic understanding of their machine learning workflow, the reasons why they choose to use auto-ml and their perceptions of auto-ml. We interviewed with 20 automated machine learning users.

We conducted the interviews from October 2019 to March 2020.The Interviews consisted three stages. First, the participants were asked to describe their job functions and their experience levels with programming and machine learning. Second, participants were asked to describe their experience in developing machine learning models and the challenges they face. Third, interviewers were asked to describe their experience in using automated machine learning tools and their perceptions of the tool(s).

Interviews were either in person (N =x) or conducted remotely (N=x). Each interview lasted for about one hour. Every participant is reimbursed with $15 for their time and insights.

We audio recorded and transcribed all interviews. Our interviews were semi-structured and centered around the following questions: [TODO: add more details depending on the findings]

- What is your current machine learning workflow with and without automated machine learning?
- How do they perceive of automated machine learning?
- If they could, what would automated machine learning users change about machine learning packages and automated machine learning tools?

4 STUDY ANALYSIS

We analyze our study data agnostic of specific auto-ml tools but we aggregated our findings accounting for the categorical differences among the tools in our analysis of usage behaviors and perceptions, for example, open source vs. commercial tools.

We engaged in an iterative and collaborative process of inductive coding to extract common themes that repeatedly came up in our data. After completing the interviews, we met weekly and discussed themes and concepts as we continued our fieldwork. We used Dedoose, an online tool for open coding, to map data onto these categories. Each of first three authors independently coded the entire data. We conducted a categorization exercise in which we physically laid out themes and relevant quotes into emerging categories. Some of our initial categories included [TODO]. Through the open coding phase, the category of [TODO...] was the most pervasive, occurring in all of our transcripts.

5 DISCUSSION: DESIGN IMPLICATIONS

6 CONCLUSION

Received April 2019; revised June 2019; accepted August 2019