Machine Learning Challenges as a Platform for Data Science Education and Public Service

Jonathan C. Stroud
stroud@umich.edu

Arya Farahi
aryaf@umich.edu

1 The Michigan Data Science Team

The Michigan Data Science Team (MDST)[1] is a student organization at the University of Michigan, Ann Arbor, which brings together students and faculty of various disciplines to collaborate on data science projects. Through these projects, MDST's goal is to provide practical educational opportunities to students while building impactful solutions to real-world problems. Many of the problems we address involve collaboration with non-profit organizations and local government, as we have found that these groups often have few resources to address their data problems on their own. Building strong, lasting relationships with these organizations is mutually beneficial, as these partnerships often result in several student projects and novel research. When developing these partnerships, we frequently use machine learning challenges as a catalyst for collaboration.

Machine learning challenges are the perfect springboard for strong partnerships, as they allow hundreds of students to effectively engage on a public service project simultaneously. During each competition, the students who are the most skilled and passionate about the project quickly rise through the ranks, and typically place highly on the competition leaderboard. We encourage these students to continue working with our sponsors, leading to more focused projects where the students work directly with collaborators from the sponsoring organization. These projects typically result in insights or tools that have great value to our community partners.

As a representative example, we examine our recent partnership with the city of Detroit, Michigan, which began with the MDST-sponsored Blight Compliance Prediction Challenge, and has resulted in several spin-off projects and publications. These ongoing projects serve as ideal examples of data science service-learning [2], and have had great value to our collaborators in Detroit. We hope that this partnership can serve as a model for using machine learning challenges for public service and data science education.

2 A Successful Partnership - Ending Urban Blight in Detroit

In recent decades, the City of Detroit, Michigan, has seen increasing numbers of investors buy up large numbers of cheap, vacant properties in the city and leave them unmaintained. This has led to a problem known as *urban blight*, where large portions of the city become uninhabitable due to excessive waste and vandalism. To combat this problem, Detroit began issuing blight violation notices in 2005, which required owners to upkeep their properties or face steep fines. Unfortunately, this program has convinced very few property owners to fix their properties, and fewer than 7% of the remaining owners have chosen to pay the resulting fines. In February 2017, MDST initiated a collaboration with the City of Detroit to provide city officials with evidence-based recommendations to improve compliance.

To kick off this partnership, MDST hosted a data science challenge centered around the blight violation data. The competition, designed in collaboration with members of the Detroit Department Department of Innovation and Technology and Department of Administrative Hearings, was specifically built to raise awareness about the problem and promote collaboration among students. The competition had two tracks:

[1] http://midas.umich.edu/mdst/

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

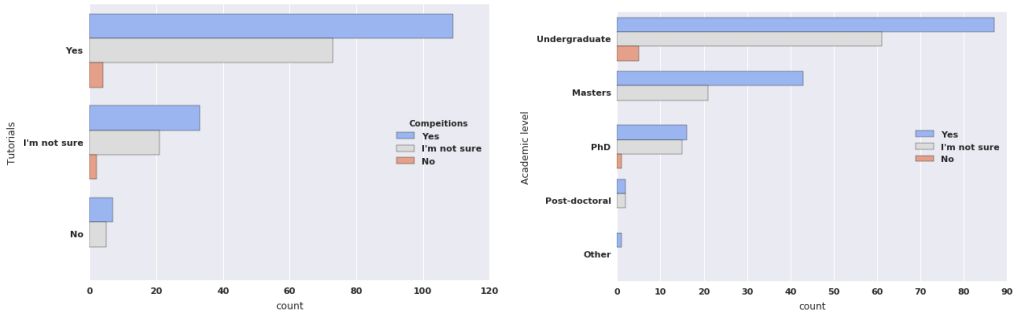

Figure 1: Selection of MDST sign-up results. Left: Students response on participating in MDST competitions and Tutorials. Right: Students response on participating in MDST competitions, grouped by their academic level.

1. **Visualization challenge**. Participants prepared data visualizations that provided novel insights about blight ticket compliance, and were judged by city officials.
2. **Prediction challenge**. Participants built models to predict whether or not a given blight violation ticket would be paid, and were judged in terms of accuracy.

Contestants were only provided a small, cleaned dataset containing blight ticket information, but were encouraged to incorporate public data from the Detroit Open Data Portal[2] into their models. In total, MDST allocated 500\$ in cash prizes for the two tracks, including a prize for the best use of external data.

The prediction challenge was hosted on Kaggle inClass[3], and attracted 39 undergraduate, graduate, and post-doctoral participants from the University of Michigan.

The visualization challenge was hosted at an event on the University of Michigan campus, and both student competitors and officials from the City of Detroit were invited to attend. The winning entry included an interactive data map that compared factors within the blight ticket dataset with factors selected from demographic data on the Detroit Open Data Portal. The event allowed student participants and city officials to interact directly and discuss aspects of the project not considered during the prediction challenge. This conversation fostered further collaboration, and lead to a paper presented at the Data Science for Social Good conference [1].

3 Machine Learning Challenges for Student Engagement

The membership of MDST has grown dramatically during the past two years, rising from just 12 students in 2015 to more than 250 students in Fall 2017. MDST is interdisciplinary, and includes students from Engineering, Statistics, Information, Physical Sciences, and several other disciplines. We find that machine learning challenges are an effective way to engage a large, interdisciplinary group like MDST, as it provides an easy avenue for students to try out techniques that they learned from coursework in a low-stakes environment.

Based on a survey of 52 MDST Members in 2016, we learned that *education* is the most important motivation for new members [3]. No respondents marked that winning prizes in competitions was a primary motivating factor, implying that education and applications are much more important than large prizes in competitions. We again surveyed our membership to find out about their attitude toward data science competitions (see Figure 1). This survey has 254 responses in total. The majority of students, 59% responded that they were interested in joining MDST sponsored competitions, while 38% were not sure and only 2% did not want to participate. This trend is observed for all academic levels and is highly correlated with student involvement with other MDST initiatives, such as the MDST tutorial series.

With the knowledge that competitions are of high interest to such a large number of students, we hope to continue using this model to engage students on valuable public service projects.

[2] `https://data.detroitmi.gov/`

[3] `https://www.kaggle.com/c/detroit-blight-ticket-compliance`

Acknowledgments

The authors acknowledge the support of the Michigan Institute for Data Science (MIDAS).

References

[1] Cell, Allie, Bhavika Reddy Jalli, Adam Rauh, Xinyu Tan, Jared Webb, Joshua Bochu, Arya Farahi, Danai Koutra, Jonathan Stroud, and Colin Tan (2017). Understanding Blight Ticket Compliance in Detroit. *Data Science for Social Good 2017*.

[2] Furco, Andrew. (1996) Service-learning: A balanced approach to experiential education.

[3] Stroud, Jonathan, Alex Chojnacki, and Jacob Abernethy (2016). The Michigan Data Science Team: A Student Organization for Machine Learning Challenges. *NIPS Workshop on Challenges in Machine Learning 2016*.