The Michigan Data Science Team: A Student Organization for Machine Learning Challenges

Jonathan Stroud
stroud@umich.edu

Alex Chojnacki
thealex@umich.edu

Jacob Abernethy
jabernet@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 hosts and participates in machine learning prediction challenges. When MDST was founded in Fall 2015, it was a small group of a dozen students that entered into machine learning challenges. In the past year, it has grown to include over 50 active members from many academic backgrounds. To support our growing membership, we now also organize a student-taught data science tutorial series and work on academic research and consulting projects. Our most impactful project has been on recovery from the Flint Water Crisis, in partnership with the University of Michigan, Flint, and Google.org. This project led to a publication in the Bloomberg Data For Good Exchange in September. We have also launched a prediction challenge centered around this project.

Over the course of the last year, we have learned a lot about how to run a successful machine learning student organization. We provide some results from a survey of our membership which provides some insight into student preferences. Based on what we have learned, we provide a number of "design principles" that will hopefully assist anyone seeking to emulate our success.

2 Survey Results

We surveyed our membership to find out what motivating factors caused them to join and continue to participate in MDST (see Figure 1). This survey had 52 responses in total. We find that education is the most important motivation for new members, with over 65% of respondents answering that "I wanted to learn more about machine learning/data science techniques" was their primary motivation for joining MDST. The second-highest motivating factor was "I wanted to apply what I had learned in class to real-world problems". 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 also asked students how much the prize pools motivated them when working specifically on machine learning challenges. Over 28% of respondents said that prize pools motivated them "not at all", and only 13% claimed that prize pools motivated them "very strongly". Based on this information, we know that investing monetary resources into large cash prizes is much less important than investing in educational opportunities and interesting application areas.

3 Design Principles

In this section we describe some of the key principles that we use to design the organizational structure of MDST. We believe that these principles can be used to establish successful student organizations at other universities.

Student Leadership

It is crucial that the group remains primarily a student-run organization. This prevents MDST from feeling like classwork, and provides opportunities for students to learn valuable leadership skills.

---

[1] mdst.eecs.umich.edu

(a) "What was your primary reason for joining MDST? Choose the best answer."

(b) "When you compete in MDST prediction challenges, how much does the prize pool motivate you to participate?"
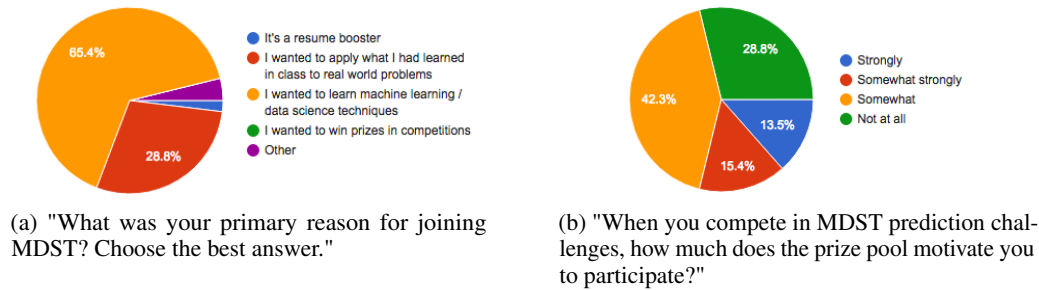
Figure 1: Selection of MDST survey results.

While we do include faculty advisers in our leadership, there are a number of ways this principle affects our daily operations.

- Weekly meetings and email communications are run by student leaders.
- Students choose their own projects and schedules.
- Projects are matched with experienced student mentors.

Having student leaders run weekly meetings is crucial because it conveys the fact that any student could become an engaged and active leader in the group during their time as a student.

Incentivize Engagement

Students are primarily motivated by learning opportunities, not financial gain. For this reason, we invest more resources in providing valuable educational incentives, rather than cash prizes, for competitions. This can come in many forms.

- Competition winners write educational blog posts about their solutions.
- Students have opportunities to co-author academic papers.
- Challenges are selected from a diverse range of application areas.

Typically, cash prizes for our competitions total no more than $1000 for the top three teams.

Education and Applications in Parallel

Many students that join MDST are relatively inexperienced with data science, but are eager to get involved right away. On the other hand, much of MDST's heavily-active membership has extensive data science experience and seek to apply novel techniques to challenging problems. To balance the needs of these two groups, we provide new students with two entry points for joining MDST.

For the new and inexperienced students, we provide a comprehensive tutorial series which, beginning with the first lesson, teaches practical application of machine learning techniques. From here, eager students naturally gravitate towards our project teams. For junior members of MDST, we set aside practical mini-projects relating to our larger ongoing projects. These mini-projects pair a group of students and a mentor with a problem to be solved. For the experienced students, we hold weekly project meetings, where we review recent research and work together on prediction challenges and our larger ongoing projects. This structure has been very effective in keeping both groups engaged, providing opportunity to progress and advance for all skill levels of our membership.

4 Future Research

Curiously, while not many students claimed that the prize pool was a significant motivating factor in machine learning challenges, no respondents answered that they were "very confident" that were capable of winning a prize in an MDST challenge. Perhaps this is the reason very few students see prizes as a motivating factor – they do not believe that these prizes are within reach. In the future, we hope to make these prizes more accessible to all students to encourage further engagement.