UCD4IDS Program Survey

Results from Faculty and Student Perspectives

Executive Summary

The UC Davis TETRAPODS Institute of Data Science (UCD4IDS) is supported by the National Science Foundation (NSF) to unify data science research and education across four disciplines: Computer Science, Electrical & Computer Engineering, Mathematics, and Statistics. The primary goal of UCD4IDS is to identify and address cross-interdepartmental barriers and encourage interdisciplinary research collaborations among faculty members, postdocs, and graduate students. To this end, UCD4IDS organized various program activities including but not limited to, round-table discussions, quarterly colloquia, and annual workshops to foster an interdisciplinary data science community at UC Davis.

UCD4IDS partnered with the Institute for Social Research (ISR) to evaluate the impact of UCD4IDS program activities and gain a better understanding of ways the program could be improved in the future. Consequently, program surveys were created and distributed to both faculty[1] (N=39) and student[2] (N=49) participants between December 3, 2022 – January 6, 2023. However, due to low response rates, the deadline was extended to January 20, 2023. At the end of data collection, response rates for faculty members reached 79 percent and student response rates reached 78 percent.

Key Insights & Recommendations

Seminar & Colloquia | Overall, more than 90 percent of respondents rated the quality of seminars as good, very good, or excellent. Similarly, more than 80 percent of respondents rated the quality of colloquia positively[3]. However, a majority of respondents were not aware of seminar recordings and PowerPoint presentations available online. As a result, students recommended consolidating information regarding program activities in one place. In doing so, respondents may be more aware of resources and keep up-to-date with relevant program information.

Hardware & GPU | More than 70 percent of faculty and student respondents who have established an account on the GPU system found it useful. At the same time, faculty respondents reported several difficulties using this system. Faculty and student respondents emphasized the need for additional resources on how to use the GPU system.

Collaborations | More than 80 percent of faculty respondents felt enough opportunities were given to collaborate compared to student respondents (51%). Coordinating communication between collaborators emerged as the one major difficulty experienced by both faculty and students.

Software & Website Issues | Although a majority of respondents are aware of the GitHub website, fewer faculty and student respondents have utilized the website to download codes. Recommendations include clarity between various relevant websites and the need for a website coordinator.

Future of Data Science | While the majority of respondents were in favor of establishing a graduate group in data science, respondents also expressed concerns regarding various changes. Importantly, faculty and student respondents suggested ways to improve data science research, education, and activities which include increasing the number of professors and courses available, creating data science competitions to engage students, and providing more opportunities to build industry connections.

[1] Faculty includes the lead PI, co PI, faculty, postdocs, alumni faculty, and alumni postdocs.
[2] Student includes graduate students and student alumni.
[3] Positively represents respondents who indicated 'good', 'very good' or 'excellent'.

Introduction

The UC Davis TETRAPODS Institute of Data Science (UCD4IDS) aims to build interdisciplinary research collaborations among faculty members, post-docs, and graduate students across four departments: Computer Science, Electrical & Computer Engineering, Mathematics, and Statistics. In order to evaluate their current programs, the Institute for Social Research (ISR) is working with UCD4IDS to serve as an external evaluator.

The UCD4IDS Program Survey was created to gain a deeper understanding of how program activities have impacted the UCD4IDS community. The following report documents the results and themes that have emerged from survey data collected from faculty respondents, and student respondents, as well as an overall look into the program activities.

Seminars & Colloquia

Seminar Attendance

Respondents were asked how often they attended various seminars offered at UC Davis as well as seminars organized by other institutions. Results demonstrate that a higher percentage of faculty respondents (75%) indicated 'sometimes', 'often', or 'always' attending the MADDD Seminar Series compared to student respondents (61%). Similarly, a higher percentage of faculty respondents (65%) indicated 'sometimes', 'often', or 'always' attending the Statistics Seminar Series compared to student respondents (48%). However, the opposite is true for zoom-based seminars organized by outside institutions; 60% of student respondents attended zoom-based seminars hosted by outside organizations compared to 43% of faculty respondents.

Table 1 | How often have you attended…

		Faculty	Student	Overall
…MADDD Seminar Series	Never	2 (7%)	1 (3%)	3 (4%)
	Rarely	6 (19%)	14 (37%)	20 (29%)
	Sometimes	16 (52%)	14 (37%)	30 (44%)
	Often	5 (16%)	9 (24%)	14 (20%)
	Always	2 (7%)	0 (0%)	2 (3%)
	Total	31 (100%)	38 (100%)	69 (100%)
…Statistics Seminar Series	Never	3 (10%)	4 (11%)	7 (10%)
	Rarely	8 (26%)	16 (42%)	24 (35%)
	Sometimes	9 (29%)	11 (29%)	20 (29%)
	Often	8 (26%)	6 (16%)	14 (20%)
	Always	3 (10%)	1 (3%)	4 (6%)
	Total	31 (100%)	38 (100%)	69 (100%)
…zoom-based seminar series organized by other institutions or groups[4]?	Never	9 (29%)	5 (14%)	14 (21%)
	Rarely	9 (29%)	10 (27%)	19 (28%)
	Sometimes	11 (36%)	18 (49%)	29 (43%)
	Often	2 (7%)	4 (11%)	6 (9%)
	Always	0 (0%)	0 (0%)	0 (0%)
	Total	31 (100%)	37 (100%)	68 (100%)

[4] Example seminar series include: One World Seminar Series on the Mathematics of Information, Data, and Signals (1W-MINDS); One World Seminar Series on the Mathematics of Machine Learning (1W-ML), etc.

Colloquia Attendance

In terms of colloquia attendance, more than three-fourths (87%) of faculty respondents attended the appropriate colloquia whereas, a little more than half (58%) of student respondents attended the appropriate colloquia. Overall, a majority of respondents (71%) have attended the appropriate colloquia.

Table 2 | Colloquia Attendance

		Faculty	Student	Overall
Have you attended the appropriate colloquia (e.g., the joint Math/Stat colloquium; ECE colloquium; etc.)?	Yes	27 (87%)	22 (58%)	49 (71%)
	No	4 (13%)	16 (42%)	20 (29%)
	Total	31 (100%)	38 (100%)	69 (100%)

Frequency of Seminars & Colloquia

Respondents were asked their opinion about the frequency of seminars and colloquia offered at UC Davis. The majority of faculty respondents (84%) and student respondents (78%) indicated the number of seminars at UC Davis were adequate. Similar trends were found for the number of data-science oriented colloquia held at UC Davis, in which 90% of faculty respondents indicated the number of colloquia held at UC Davis was adequate compared to 70% of student respondents. None of the faculty or student respondents expressed there were too many colloquia. Moreover, 30% of student respondents indicated a need for *more* data-science oriented colloquia held at UC Davis.

Table 3 | What do you think about the frequency of…

		Faculty	Student	Overall
… data-science oriented **seminars** held at UC Davis?	Too Many	1 (3%)	0 (0%)	1 (2%)
	Adequate	26 (84%)	29 (78%)	55 (81%)
	Needed More	4 (13%)	8 (22%)	12 (18%)
	Total	31 (100%)	37 (100%)	68 (100%)
… data-science oriented **colloquia** held at UC Davis?	Too Many	0 (0%)	0 (0%)	0 (0%)
	Adequate	28 (90%)	26 (70%)	54 (79%)
	Needed More	3 (10%)	11 (30%)	14 (21%)
	Total	31 (100%)	37 (100%)	68 (100%)

Quality of Seminars & Colloquia

As for the quality of seminars and colloquia held at UC Davis, faculty respondents rated the quality of seminars (97%) and colloquia (96%) positively[5]. Student respondents held similar positive perceptions of the quality of seminars (89%) and colloquia (83%).

Table 4 | What do you think about the quality of …

		Faculty	Student	Overall
… data-science oriented **seminars** held at UC Davis?	Poor	0 (0%)	0 (0%)	0 (0%)
	Fair	1 (3%)	4 (11%)	5 (7%)
	Good	7 (23%)	14 (38%)	21 (31%)
	Very Good	18 (58%)	16 (43%)	34 (50%)
	Excellent	5 (16%)	3 (8%)	8 (12%)
	Total	31 (100%)	37 (100%)	68 (100%)
… data-science oriented **colloquia** held at UC Davis?	Poor	1 (3%)	0 (0%)	1 (2%)
	Fair	0 (0%)	6 (17%)	6 (9%)
	Good	7 (23%)	18 (50%)	25 (38%)
	Very Good	13 (43%)	9 (25%)	22 (33%)
	Excellent	9 (30%)	3 (8%)	12 (18%)
	Total	30 (100%)	36 (100%)	66 (100%)

Access to Seminars & Colloquia

The UCD4IDS program offers PowerPoint presentations of seminars and colloquia on its website. While a little more than half (58%) of faculty members were aware of this, a majority (61%) of student respondents were not aware of this resource. Of those who were aware of the PowerPoint presentations, less than half (39%) of faculty members accessed this resource whereas 67% of student respondents indicated visiting and viewing the PowerPoint presentations available.

Table 5 | Awareness and Usage of PowerPoint presentations of seminars and colloquia

		Faculty	Student	Overall
Are you aware that PowerPoint presentations from many of these seminars/colloquia are available at the following website: https://www.math.ucdavis.edu/~saito/ucd4ids/?	Yes	18 (58%)	15 (40%)	33 (48%)
	No	13 (42%)	23 (61%)	36 (52%)
	Total	31 (100%)	38 (100%)	69 (100%)
Have you ever visited the website and viewed any of the PowerPoint presentations available?	Yes	7 (39%)	10 (67%)	17 (52%)
	No	11 (61%)	5 (33%)	16 (49%)
	Total	18 (100%)	15 (100%)	33 (100%)

Response to COVID-19

In response to COVID-19, the departments associated with UCD4IDS organized various zoom-based seminars. More than three-fourths of faculty (81%) and student (79%) respondents have attended zoom-based seminars. These seminars were recorded and made available on the UC Davis website. However, a lower percentage of faculty (52%) and students (45%) were aware of the recordings and access to the

[5] Positively represents respondents who indicated 'good', 'very good' or 'excellent'.

presentations. Of those who were aware, half (50%) of faculty respondents and 63% of student respondents indicated viewing at least one of these videos.

Table 6 | UC4IDS Zoom-based seminars

		Faculty	Student	Overall
In response to COVID-19, the departments associated with UCD4IDS organized various zoom-based seminars. Have you ever attended at least one of these seminars?	Yes	25 (81%)	30 (79%)	55 (80%)
	No	6 (19%)	8 (21%)	14 (20%)
	Total	31 (100%)	38 (100%)	69 (100%)
Are you aware that quite a number of zoom-based seminars were video recorded and viewable on AggieVideo using the following link: http://https://video.ucdavis.edu?	Yes	16 (52%)	17 (45%)	33 (48%)
	No	15 (48%)	21 (55%)	36 (52%)
	Total	31 (100%)	38 (100%)	69 (100%)
Have you ever watched at least one of those videos?	Yes	8 (50%)	10 (63%)	18 (56%)
	No	8 (50%)	6 (38%)	14 (44%)
	Total	16 (100%)	16 (100%)	32 (100%)

Round Table Discussions

Seminars were accompanied by roundtable discussions during Fall 2019 and Winter 2020. Participation in the roundtable discussions varied between faculty and student respondents. More faculty respondents (61%) attended/participated in roundtable discussions compared to student respondents (24%). At the same time, both faculty (90%) and student (89%) respondents agreed that the round table discussions were useful. Although the roundtable discussion minutes are available on the UCD4IDS website, less than a third of faculty (19%) and students (27%) respondents accessed this resource.

Table 7 | Awareness and usefulness of roundtable discussions

		Faculty	Student	Overall
Have you attended/participated in the roundtable discussion associated with the seminars (during Fall 2019 and Winter 2020)?	Yes	19 (61%)	9 (24%)	28 (41%)
	No	12 (39%)	28 (76%)	40 (59%)
	Total	31 (100%)	37 (100%)	68 (100%)
Did you find those roundtable discussions informative or useful?	Yes	17 (90%)	8 (89%)	25 (89%)
	No	2 (11%)	1 (11%)	3 (11%)
	Total	19 (100%)	9 (100%)	28 (100%)
Have you ever looked at the roundtable discussion minutes that are available from our website found here: https://ucd4ids.ucdavis.edu?	Yes	6 (19%)	10 (27%)	16 (24%)
	No	25 (81%)	27 (73%)	52 (77%)
	Total	31 (100%)	37 (100%)	68 (100%)

Suggested Speakers for Future Colloquia

Generally, faculty respondents suggested that there should be colloquia for new hires in data science at least once a year. They also shared they were most interested in hearing from people that lead teams in cutting-edge machine learning firms such as Goodfellow. Other suggestions include learning from those who work with "VAE, GANs, transformers, language models," or individuals from "SV (low cost) on RL, AL, etc."

Faculty and student respondents were also asked to suggest potential speakers for future colloquia. The following list of speakers was recommended by faculty and student respondents:

- Al Hero (Michigan)
- Andrew Ng*
- Aukosh Jagannath*
- Dan Spielman (Yale)
- David Donoho (Stanford)
- Dr. Robert Lund, Chair of Statistics, UCSC
- Dustin Mixon (1W-Minds)*
- Heather Harrington (Oxford)
- Jianqing Fan*
- Judea Pearl
- Lihua Lei*
- Michael Bronstein*
- Michael Schaub*
- Peter Hall*
- Rahul Mazumder (MIT)
- Ryan Tibshirani
- Vincent Hellendoorn, CMU
- Weijie Su*
- Yang Song (OpenAI)*
- Yann LeCunn
- Yi Ma, UC Berkeley*
- Zhi Ding

Please note asterisks (*) denote speakers who were recommended by students. The only speaker who was recommended by both student respondents and faculty respondents was Yi Ma.

Improvements for Seminars, Colloquia, and/or Roundtables

Several faculty respondents expressed appreciation for the current available seminars; one faculty respondent suggested access to seminars via Zoom recordings was particularly appreciated as they allow for more flexibility for those who would like to attend but are no longer working at UC Davis. Another faculty respondent questioned what were the primary goals of these events. They proposed that while bringing speakers who have a visible presence on papers and applications (e.g. ChatGPT) may increase event attendance, they were uncertain as to whether increased attendance was the goal of the program. Other suggestions include arranging meetings between speakers and students as well as increasing the number of in-person seminars and colloquia.

On the other hand, student respondents expressed the need for invited speakers to provide more background information and "big picture" research to accommodate diverse audiences. In doing so, audiences who come from a variety of different fields would be less likely to "get lost" during the presentations. Logistically, another student noted hosting the seminars in a bigger space could be beneficial. Lastly, a student respondent stated the need for a unified website in which users could access announcements, available slides, and videos in one designated place. They further suggested this could be a portal on the departmental website.

Overall, faculty and student respondents offered varied strategies to improve current seminars, colloquia, and roundtable discussions.

Hardware & GPUs

While a majority of faculty (87%) and student (70%) respondents are aware of the GPU system provided by the campus, a smaller percentage of faculty (50%) and student (38%) respondents have actually established an account and utilized the system. Of those who have used the GPU system, approximately three-fourths of faculty respondents (73%) and student respondents (71%) have found this system useful.

Table 8 |Hardware & GPUs

		Faculty	Student	Overall
Are you aware that you could use the GPU system provided by the campus called the High-Performance Computing (HPC) Core facility?	Yes	26 (87%)	26 (70%)	52 (78%)
	No	4 (13%)	11 (30%)	15 (22%)
	Total	30 (100%)	37 (100%)	67 (100%)
Have you or your team members established an account on that GPU system and run your jobs?	Yes	15 (50%)	14 (38%)	29 (43%)
	No	15 (50%)	23 (62%)	38 (57%)
	Total	30 (100%)	37 (100%)	67 (100%)
Have you found the use of our GPU system useful?	Yes	11 (73%)	10 (71%)	21 (72%)
	No	4 (27%)	4 (29%)	8 (28%)
	Total	15 (100%)	14 (100%)	29 (100%)

Hardware & GPU Improvements

Although some faculty respondents are thankful for the availability of GPUs, many faculty respondents suggested the current hardware and GPUs need improvements (please see Appendix A for full list of responses). Several faculty members reported experiencing issues "running a job for a long time" and suggested it is "very cumbersome to use". From a campus-wide infrastructure perspective, faculty respondents stated the HPC is fragmented across different groups and colleges, for example, HPC1 is GPU based but HPC2 is not. Faculty respondents also noted a lack of campus-wide HPC support. They urge UC Davis's administration to pay more attention to this group and the facilities. Moreover, faculty respondents emphasize the need to improve communication with the computing administrative team in order to make informed quality decisions with appropriate faculty consultation. Largely, the lack of support and resources to use GPUs causes problems for users. Faculty respondents explained the need for more nodes to meet increasing demands and additional support for students who may be new to this system of computing. One suggestion is to offer short courses on how to use the GPUs in addition to providing more information to the students.

Student perspectives echoed many of the same sentiments in which the difficulty of usage, the need for instruction and training on how to use the GPUs, as well as the need for GPUs with more memory are all barriers that prevent increased usage and usefulness. Student respondents requested tutorials, user interface instructions, trainings, or information sessions to introduce the high-performance GPU platform and demonstrate how to access this resource.

Collaborations

The degree to which faculty and student respondents felt enough opportunities have been given to collaborate varied. A majority of faculty respondents (83%) believed enough opportunities have been given to collaborate whereas student respondents were more divided in their responses. A little more than half of student respondents (51%) believed enough opportunities have been given to collaborate.

Table 9 | Opportunities for Collaborations

		Faculty	Student	Overall
Do you feel that enough opportunities have been given (via seminars, email communications, etc.) to collaborate?	Yes	25 (83%)	18 (51%)	43 (66%)
	No	5 (17%)	17 (49%)	22 (34%)
	Total	30 (100%)	35 (100%)	65 (100%)

Strategies to Meet Collaborators

While most student respondents developed collaborations through the help of their supervisors, advisors, and connections within their departments, faculty members expressed more diverse avenues of collaboration. When asked how faculty respondents got to know their collaborators, a majority of faculty respondents mentioned events such as seminars, workshops, and conferences as a main mode of connection. Another strategy faculty respondents used to get to know their collaborators is through informal interactions with colleagues and graduate students.

Collaborations within UC Davis and outside of UC Davis

Only a few student respondents provided information regarding their collaborations both within UC Davis (N=2) and outside of UC Davis (N=3). Student respondents indicated most of these collaborations were indeed interdisciplinary (please see Appendix B for a full list of student collaborations).

Response rates were higher for faculty respondents which 12 faculty respondents indicated having collaborations within UC Davis and 14 faculty respondents indicated having collaborations outside of UC Davis. A majority of faculty indicated their collaborations were interdisciplinary (please see Appendix C for a full list of faculty collaborations)

Major Difficulties in Conducting Data Scientific Collaborations

Despite the collaborations that occur both within and outside of UC Davis, faculty respondents also shared some of the difficulties associated with conducting data scientific collaborations. One theme that emerged is the difficulty of communication across disciplines where different theories or specific languages are used in each domain. In addition to working with different experts in different fields, coordinating communication across different time zones can also be hard to navigate. Communication difficulties are exacerbated by COVID-19 restrictions where conducting in-person meetings or coordinating travel are more complicated. Faculty respondents also noted the difficult nature of discussing mathematical proofs as well as the logistics of sharing data across institutions. Related to data, one faculty respondent also explained how collecting, managing, and preprocessing data is a strenuous process because students do not show interest in these aspects of the research process. With these difficulties in mind, another faculty member simply stated the lack of time to conduct collaborations.

From the student perspective, they also stated difficulty in having in-person discussions and noted that some researchers do not like to respond to emails or share original codes. Additionally, resources for GPUs that can handle language modeling are limited.

Software & Website Issues

Faculty and student respondents were also asked whether they were aware of numerous websites available for their use. A majority of faculty (87%) and student (95%) respondents were aware of the

UCD4IDS website and have visited the website (faculty = 92%, student = 91%). At the same time, when asked whether they were aware of the UCD4IDS GitHub repository, the percentage of faculty and student awareness declined. Forty-three percent of faculty respondents were aware of the repository whereas 40% of student respondents were aware of the repository. For those who were aware, more than half of faculty respondents (62%) visited the website and more than three-fourths (79%) of student respondents visited the website. However, when asked if respondents ever downloaded any codes listed on the GitHub website, a majority of faculty (77%) and student (71%) respondents have not utilized this resource.

Table 10 | Software & Website Awareness and Usage

		Faculty	Student	Overall
Are you aware that the UCD4IDS has its own website: https://ucd4ids.ucdavis.edu?	Yes	26 (87%)	35 (95%)	61 (91%)
	No	4 (13%)	2 (5%)	6 (9%)
	Total	30 (100%)	37 (100%)	67 (100%)
Have you ever visited that website?	Yes	24 (92%)	31 (91%)	55 (92%)
	No	2 (8%)	3 (9%)	5 (8%)
	Total	26 (100%)	34 (100%)	60 (100%)
Are you aware that the UCD4IDS has its GitHub repository: https://github.com/UCD4IDS?	Yes	13 (43%)	14 (40%)	27 (42%)
	No	17 (57%)	21 (60%)	38 (59%)
	Total	30 (100%)	35 (100%)	65 (100%)
Have you ever visited that GitHub website?	Yes	8 (62%)	11 (79%)	19 (70%)
	No	5 (39%)	3 (21%)	8 (30%)
	Total	13 (100%)	14 (100%)	27 (100%)
Have you ever downloaded codes listed in that GitHub website?	Yes	3 (23%)	4 (29%)	7 (26%)
	No	10 (77%)	10 (71%)	20 (74%)
	Total	13 (100%)	14 (100%)	27 (100%)
Would you be willing to contribute your own codes to disseminate through our GitHub website?	Yes	26 (90%)	27 (77%)	53 (83%)
	No	3 (10%)	8 (23%)	11 (17%)
	Total	29 (100%)	35 (100%)	64 (100%)

Reasons that Prevent Contributions to Coding on GitHub

Ninety percent of faculty respondents indicated willingness to contribute their own codes through the GitHub website whereas 77 percent of student respondents indicated willingness to contribute their own codes. In spite of that, faculty and student respondents also expressed reasons that would prevent them from sharing their codes.

Student respondents stated the following reasons that prevent them from contributing codes to the GitHub website include: easy to do on your own GitHub, graduated from UC Davis, and current research is not heavily coding related. On the other hand, faculty respondents suggest working for industries that often times pay for their codes. Consequently, there isn't much incentive to share their codes for free on the UCD4IDS repository.

Website Suggestions

Overall, there weren't many suggestions from either students or faculty respondents. However, student respondents expressed some confusion regarding the UCD4IDS website compared to the departmental website. Faculty respondents suggest the need for a website coordinator or manager who could update the contents of the website frequently.

Suggestions on How to Disseminate One's Software Packages

Most suggestions came from faculty respondents who expressed the need for common standards and practices to be adopted. For example, adopt a release schedule, good documentation standards, and well-developed demos. In terms of packaging software, some faculty suggest putting packages on GitHub while others believe they should be published on R. Although only one student response was captured, they echoed the need for examples or demos on how to use the packages.

Future of Data Science on Our Campus

Respondents were asked about several changes that could occur to enhance the future of data science at the UC Davis campus. First, respondents were asked if they were aware of the newly established Undergraduate Data Science Major in the Fall of 2022. An overwhelming majority (93%) of faculty respondents were aware of this change, compared to less than half (44%) of student respondents.

Table 11 | Awareness of UC Davis Undergraduate Data Science Majors

		Faculty	Student	Overall
Are you aware that UC Davis started Undergraduate Data Science Majors in Fall 2022?	Yes	28 (93%)	15 (44%)	43 (67%)
	No	2 (7%)	19 (56%)	21 (33%)
	Total	30 (100%)	34 (100%)	64 (100%)

Graduate Group in Data Science

Secondly, respondents were asked whether it would be a good idea to establish a Graduate Group in Data Science (GGDS) for graduate level students. Although a majority of students indicated it was a good idea to establish the GGDS, the degree to which faculty respondents believed it was a good idea was less than that of students. Eighty-six percent of student respondents were in favor of the establishment, while 58 percent of faculty believed this was a good idea.

Table 12 | Do you think it is a good idea to establish the…

		Faculty	Student	Overall
Graduate Group in Data Science (GGDS) for graduate level students?	Yes	15 (58%)	25 (86%)	40 (73%)
	No	11 (42%)	4 (14%)	15 (27%)
	Total	26 (100%)	29 (100%)	55 (100%)

Faculty proponents of the GGDS state the primary benefit stems from encouraging interdisciplinary work. They believe the establishment of GGDS could increase interactions and faculty collaborations. Student proponents held similar beliefs. They highlighted various benefits including connecting students with similar interests, usefulness for those who want to go into the industry, increasing desirability in the market, enhancing collaboration and assisting in fast-paced learning, and creating more opportunities for students. Student respondents believe this is a necessary step due to the current needs of the industry.

While a majority of respondents believed the GGDS was a good idea, faculty and student respondents also raised concerns about the GGDS. Funding issues and overlap with current programs are primary concerns from faculty respondents. Student respondents also shared similar concerns related to overlap with GGAM. Opponents (both faculty and students) of the GGDS establishment are unclear why the

establishment of GGDS is necessary and how it could be beneficial (please see Appendix D for full list of comments regarding the GGDS).

Designated Emphasis on Data Science

In response to the idea of establishing the Designated Emphasis on data science[6], faculty members were more hesitant (please see Appendix E for full list of comments). Some faculty members expressed positive or neutral feelings, believing this could increase cross-disciplinary work. Others indicated they were in favor of the idea but qualified their answers with additional concerns. For example, while some faculty respondents believe it is a "good idea" they also felt more clarification and funding would be needed to make a meaningful impact. While some faculty respondents expressed it would be a better idea than the graduate groups, other faculty respondents have noted a designated emphasis has not worked well in other programs. Consequently, they were not supportive of establishing a designated emphasis.

Similarly, student responses were also divided. Some student respondents shared it would be a "nice idea", however, other students raised skepticism. For example, some students believe that the establishment of an emphasis would require students to take enough credits related to data science, however, not many courses in their department (math/stats) would fit this criterion. Additionally, students expressed the difficulty of enrolling in data-related courses if they are not in the department. Some student respondents believe this would be a weaker proposal than the GGDS or that they serve the same purpose. Students also suggest more thought is needed to implement this change.

Getting more Resources and Support from the Administration

In terms of getting more resources from the administration, faculty respondents had more suggestions compared to student respondents (please see Appendix F for full list of responses). Faculty respondents state that the growing interest from students wanting to enter data science programs and demand for data scientists is evidence that increased support from university staff and faculty is needed. They suggested these growing demands warrant an increase in the number of courses offered. Furthermore, faculty urge UC Davis administration to pay more attention and be involved with campus-wide data science activities. Faculty respondents suggested one way to facilitate support is to hire a coordinator who could dedicate 100% of their time to campus-wide data science activities. Similarly, another faculty respondent recommended hiring a communications specialist.

Other suggestions include providing summer classes, undergraduate research, or regular classes in applied data science, focusing on interdisciplinary recruitment of faculty members with positions allocated to different departments with a data science emphasis, and providing service/consulting in grants with an associated position to coordinate groups and grants.

Student respondents suggest promoting the value of data science in various industries that can enhance the connection with companies outside academia. Another suggestion is to increase travel funding and summer school opportunities.

[6] The Designated Emphasis on Data Science can be added to existing PhDs.

Improving Data Science Research on Campus

In order to improve data science research on campus, faculty respondents suggested varied opinions (please see Appendix G for full list of responses) that include better high-performance computing (HPC) support, and more support for the Center for Data Science and Artificial Intelligence Research (CeDAR).

Faculty respondents also expressed a sense of disconnectedness and the need to "bring people together." Some respondents encouraged the facilitation of shared mentoring of students as well as the need to hire more faculty that specialize in theory of deep learning and reinforcement learning. In terms of class content, they suggest it is important for students to learn how to use existing tools to solve new problems as this is reflective of the data science industry.

Importantly, faculty respondents highlighted the need to recognize that interdisciplinary work is difficult and under-appreciated. Other suggestions also include creating a data science symposium for graduate students to compete in competitions that offer rewards and increase the interaction with industry and tech companies to evaluate the needs in the field.

Student responses revealed similar suggestions in which more collaboration with the industry where real-world data is available was important to students. Additionally, students recommended research-oriented courses that could prepare students to be at the "frontier" of the data science research field.

Students also agree that more data science faculty are needed, noting the loss of several machine learning oriented professors in the computer science department. Other suggestions include providing funding for students to attend seminars, conferences, and summer schools that may be relevant to their areas of focus. Student respondents also value interdisciplinary research projects that have influence in other communities.

Improving Data Science Education on Campus

Faculty respondents were asked if they had any suggestions to improve data science education on campus (please see Appendix H for full list of responses). Faculty encouraged events where students can present their research or discuss graduate level courses and create data science competitions (such as "Kaggle") to encourage student participation. Additionally, faculty respondents suggested hiring more faculty, creating more courses, smaller course enrollments, and increasing the frequency of courses. Courses should also cover relevant topics like sprint planning, open-source software development, and modern machine learning tools (e.g. Torch, Pytorch-lightning, DVC, etc.) One faculty respondent also encouraged early recruitment at high school or earlier.

Student respondents provided similar suggestions and stressed the need for more machine learning courses, more frequent offerings of these courses, and more faculty members in machine learning. One respondent suggested including theoretical machine learning classes in the Math/Statistics department and practical machine learning courses in the Computer Science departments. Similar to faculty responses, students suggest coursework should follow emerging trends in the field and should include diverse and modernized graduate level courses.

Moreover, student respondents stated the need to enhance communication between students (both graduate and undergraduate) and professors. More hands-on courses in deep learning as well as rigorous math/statistics trainings would be appreciated in order to help students stand out in the competitive nature of the field.

Improving Data Science Activities on Campus

Faculty respondents suggested a number of different ways to improve data science activities on campus (please see Appendix I for full list of responses). Oftentimes, faculty respondents mentioned large-scale structural changes that could benefit data science activities at UC Davis. For example, creating a physical location where people could meet (e.g. creating a data science center - similar in structure to the Genome Center) or designating a general campus facility that could serve as the location for quarterly seminars instead of a departmental facility.

In thinking about how to stay competitive with other universities, faculty respondents emphasized the importance of receiving philanthropy/donations. Determining a data science niche and could be helpful to stay competitive among other data science programs at other universities. At the same time, one faculty respondent suggested that other campuses have found success by incorporating data science in every course on campus. They suggested adding data literacy requirements to all majors and all degrees could help create a data science culture on campus and increase the desire to use data science in their specialties.

Another faculty respondent recommended learning from other examples including the CMU Delphi's approach to create groups that focus on specific goals. They emphasized the benefit of creating groups of professors and students (roughly 2 professors + 10 students) to work on specific goals as a way to create meaningful change.

Additional suggestions include working with K-12 teachers to identify younger talents, creating a career fair specifically for data science, increasing pay for machine learning professors, making data from outside collaborators more important, identifying problems where machine learning is actually needed, and establishing more collaborations with industry.

Similarly, student respondents indicated wanting to build long-term relationships with relevant industries. In addition to building industry relationships, students also suggested hosting data science conferences in Davis or to hold more educational seminars that focus on introducing current active research areas. Students also showed interest in data challenges such as Kaggle, where students (both undergrad and grad) could form groups and participate in them. Aside from data science events, student respondents recommended increasing the number of data science related courses that are made available as electives.

Summary

In order to gain a better understanding of UCD4IDS program activities, faculty, and students were asked a series of questions regarding seminars & colloquia, hardware & GPUs, collaborations, software & website issues, and the future of data science in the UC Davis campus.

Faculty respondents have attended most seminars at a higher rate than student respondents except for seminars held by other organizations. Faculty respondents also attended colloquia at a higher rate than students. A majority of faculty and student respondents believe the frequency of seminars and colloquia is adequate. Furthermore, a majority of faculty and student respondents positively rated the quality of seminars and colloquia.

While most of the PowerPoint presentations of the seminars and colloquia are available online, a majority of students were not aware of this. However, students who were aware, are also more likely to

access this resource compared to faculty respondents. This could be an indication that if more students were aware of this, they would also be more likely to utilize this resource.

Attendance of roundtable discussions demonstrated the largest discrepancy between faculty attendance and student attendance. At the same time, a majority of faculty and student respondents who attended the roundtable discussions found it to be useful.

Importantly, suggestions to improve the seminars and colloquia involve bringing speakers who have a strong presence in the field and finding ways to accommodate a diverse audience with varying degrees or knowledge.

In terms of hardware & GPUs, a majority of faculty and student respondents are aware of the HPC Core facility, data demonstrates a difference in usage where faculty respondents are more likely to use the system compared to student respondents. Furthermore, respondents highlighted issues with the system that deter usage. For example, many faculty respondents indicated difficulties running jobs and the lack of a cohesive campus-wide infrastructure to support the work that they do. While student respondents also stated difficulties with using the system, their main suggestion would be to include training/guides on how to use the GPUs.

While faculty respondents shared more diverse strategies to meet collaborators, student respondents indicated meeting their collaborators through their supervisors or advisors. Although collaborations are occurring within UC Davis and outside of UC Davis, faculty respondents explained the primary difficulties include finding the time to coordinate and communicate with people from varied fields, adjusting to COVID-19 restrictions, and finding students who are interested in all aspects of collaboration. Student respondents also mentioned the difficulty of communicating with faculty members who are not as responsive to their emails or seem reluctant to share original coding.

While most faculty and student respondents were aware of the UCD4IDS website, fewer respondents were aware of the Github Repository. While most respondents indicated a willingness to upload their own codes, several respondents were unclear as to the benefits of uploading their codes to the UCD4IDS repository. Students shared they could upload on their own repository whereas faculty respondents suggest paid incentives would be needed to encourage faculty.

Due to the various websites available, student respondents were confused as to the differences between UCD4IDS website and the departmental website. Faculty respondents suggest hiring a website coordinator to keep website content updated. Suggestions for dissemination of software highlight a need for standardization of the process.

Respondents were also asked their opinions regarding several changes to the structure of Data Science at UC Davis including their views on the newly established Graduate Group in Data Science and the Designated Emphasis for post-doctoral students. While opinions were varied, respondents shared both positive and negative responses and indicated more thought was needed to ensure meaningful impact.

Lastly, respondents were asked to suggest ways in which the future of data science could be improved at the UC Davis campus. Faculty and student respondents provided a wide range of improvements that include creating more industry connections, providing opportunities for students to compete and gain experience with real world data, and increasing data science capacity at UC Davis by offering more classes and creating a culture of data science.

Altogether, faculty and student perspectives demonstrated ways in which UCD4IDS program activities impacted the data science community at UC Davis and provided ways in which the program can be improved.

Appendix A

Raw responses from Hardware and GPU Improvements.

Faculty Responses (N=11)

- Campus HPC support is lacking
- difficult to run a job for a long time - we had to buy our own server.
- For many students this is a new way to compute. Short courses on how to use GPU would be valuable.
- HPC1 is GPU based but HPC2 is not. The computing admin team is not technically competent and makes poor decisions without proper consultation.
- I am OK with the current GPU status. Maybe more nodes should be added in future to meet increasing demands.
- It seems like the campus HPC infrastructure is currently quite fragmented across different groups and colleges and hard to buy in to
- no complaints
- Nothing specific, thanks for making facility available.
- Sending detailed information, specially to grad students would be helpful
- The campus-wide HPC is currently not organized well. The UCD administration needs to pay more attention to this group and facilities.
- Very cumbersome to use, often not working

Student Responses (N-5)

- Again, it might be helpful to create a portal with UI and instructions on how to use
- I registered but never had a chance to use it.
- I think we need some tutorial/information session to introduce the high-performance GPU platform and how to access.
- it's hard to use
- Need better documentation and training regarding the use of GPU. Also, need GPUs with more memory such as 32/64 GB for large models.

Appendix B

Student respondents were asked to list their collaborations within and outside of UC Davis (please see table below).

Collaborations (outside UCD4IDS but within UC Davis)	Interdisciplinary (Yes/No)	Comments
Emily Morgan (linguistics), Kenji Sagae (linguistics/cs)	Yes	Yes inspecting how ML models trained on code impact linguistic studies
John Albeck	Yes	My data provider is from monocular biology
Collaborations (outside UC Davis)		
Sampling problems with Murat Erdogdu, Regularized SVGD with Bharath K Sriperumbudur and Jianfeng Lu	Missing	Missing
Vaneet Aggarwal from the Purdue University	Yes	My collaborations are with Berkeley lab and Purdue University
With Prof. Lihua Lei on consistency of spectral method on community detection	Missing	Missing

Appendix C

Faculty respondents were asked to list their collaborations within and outside of UC Davis (please see table below).

Collaborations (outside UCD4IDS but within UC Davis)	Interdisciplinary (Yes/No)	Comments
UCD Health colleagues	Yes	Yes, application of data science to health diagnosis and neuropathology analysis
SOM	Yes	Yes
Data analysis project on Cow milk production and cow diseases with Fernanda Ferreira (Veterinary School)	Yes	Yes. It involves veterinary scientists, mathematicians, and statisticians.
UC Davis School of Medicine	Yes	Yes, medicine
Nil	Yes	astronomy and statistics
David Woodruff (GSM)	Yes	Yes, trying to model various problems in Energy.
With faculty members Somen Nandi and Karen McDonald (Chemical Engineering) on plant-based protein production for space explorations; With Jie Peng (Statistics) on graphical models.	Yes	The one with Chemical Engineering faculty was interdisciplinary by its very nature since it used statistical analysis tools to formulate questions about adequacy of the experimental set up associated with scientific experiments related to space exploration.
Sean Burgess, Priya Shah	Yes	Yes, combination of experiments and mathematical modeling
Randy O'Reilly, Tim Hanks, Ben Yoo, Daniel Cox	Yes	Yes. I work at the intersection of math and neuroscience and these collaborations ranged from experimental neuroscience to computer science.

Beatriz Martinez Lopez, Yueyue Fan, Tucker Jones	Yes	All of them
Emily Morgan, Kenji Sagae (both in Linguistics)*	Yes	Yes, we are (in both cases) studying the ability of Language Models to predict human behavior as observed in direct studies and observationally in the software archives.
I am a member of AIFS, so multiple collaborations with AIFS partners (UCB, Cornell, UIUC, ARS, etc.)	Yes	all of them - they have to do with use-inspired AI, Nutrition, Food, and analysis of clinical/biological/agricultural data
Collaborations (outside UC Davis)	**Interdisciplinary (Yes/No)**	**Comments**
University of Kentucky	Yes	Intra-disciplinary - collaborating with experts in image/video processing, security, and privacy
Synthetic aperture sonar data analysis with the Naval Surface Warfare Center, Panama City, FL	Yes	Yes. It involves mathematician, electrical engineers, and physicists.
Center for Astrophysics at Harvard	Yes	astronomy and statistics
Ecole Polytechnique France, TU Berlin Germany, Oxford Univ. UK	Yes	Yes, collaborators in Oxford are mostly interested in biological applications.
With Tomoko Matsuo (University of Colorado, Boulder) on high-dimensional regression for modeling earth's magnetic fields. With Sanjay Chaudhuri (National University of Singapore) on analysis of the dynamics of the covid pandemic and possible mitigation strategies that take into account their economic impact.	Yes	Both collaborations listed were interdisciplinary since they required utilizing statistical modeling and inference techniques for dealing with physical and epidemiological phenomena, respectively.
Radmila Sazdanovic (NC State), Wenqin Luo (U Penn)	Yes	Yes, combination of data science and experiments or data science and math
Dr. Rebecca Killick, Lancaster University, UK	Yes	Statistical application in climatology
UCSD: Alex Cloninger, Scott Mahan, RPI: Rongjie Lai, UDel: Nikolas Schonsheck	No	No
Kaiyi Ji	No	No, this collaborator has similar background as mine.
Kaiyi Ji (U of Buffalo), Lingzhou Xue (Penn State)	Yes	computer science, statistics, optimization.
Collaborations within Amazon	Yes	Yes, related to Amazon Search, also I have worked with CMU Delphi
Baishakhi Ray, Columbia Univ., Earl Barr, University college London, CHris Bird, Microsoft Research	No	No. All within CS
National Chung-Hsing University Taiwan, University of Gottingen Germany	Yes	Collaboration with University of Gottingen is interdisciplinary.
Gilles Blanchard, Frederic Chazal, Paris, France; Alexander Kreiss, Leipzig, Germany	No	No.

Appendix D

The table below displays all comments related to the establishment of the Graduate Group in Data Science (GGDS).

Proponents of GGDS

Faculty (N=15)

- Collaborations among granduates definitely are needed.
- Could be good to promote inter-disciplinary work
- Having this graduate group could further facilitate collaborations among faculty in this area.
- I guess it would further enhance interactions on campus, and provide more options for the students; however a thorough discussion about this needs to take place first with the many stakeholders (e.g. Appl. Math Graduate Group; Statistics Graduate Program, CS, etc.)
- I think a multi-department program would make us more attractive to graduate applications.
- I'm not sure. Could be attractive to students but I worry that a grad group in data science would try to cover too many disciplines and not provide sufficient core methods training
- It will attract more students
- It would be more appropriate than using GGAM for this field of study
- It's important to have a local community of people to talk to about reseach and new papers.
- Not essential, but none of the curricula from existing GGs fit the needs of data science.
- The field is now diverse enough to have a specific focus - what will be the home department though?
- Would be an important subject. The risk is that it becomes a disguised program coding major.

Student (N=25)

- Data Science is getting more and more popular. GGDS would connect students with similar interests.
- Good to concentrate on a particular area and spend resources on it
- I think DS is of interdisciplinary nature and should be a joint work with stat and CS. The only concern is that this might overlap with GGAM
- I think it will be really useful for folks wanting to go to Industry
- I would love to see data centric cross interactions with other students at UCD
- It's necessary due to the need from industry and the interest from students
- It's desired by the market
- More opportunities for interested students
- Students need this major
- To help collaboration and learn various aspects of this fast-pace area
- Will be more like a trade school with production-level data engineering, advanced statistics, lot of familiarity with DataFrames and similar tools

Opponents of GGDS

Faculty (N=11)

- Concern about the additional administration cost and dillution of faculty advior time/energy
- Eventually yes. But right now there is much overlap with GGAM. The benefit with more administrative work is not large enough.
- Graduate groups are poorly funded. The students are better off being in a departmental program.
- Hard to see how this would expand existing research opportunities, but would be a big administrative timesink
- It interferes many existing graduate programs and groups too much.
- The graduate groups have tended to be poorly funded, unless they are very closely tied to specific departments.
- There are many graduate groups that address data science in different ways
- Too much fragmentation
- Unclear why necessary

Student (N=4)

- Just personally feel that the statistics and applied math grad student group already have much overlap with data science research.
- Unnecessary

Appendix E

The table below displays all comments related to the establishment of the Designated Emphasis on Data Science.

Proponents of the Designated Emphasis on Data Science

Faculty (N=16)

- Better than establishing a GG in Data Science
- CS and Stat
- Designated Emphasis on Data Science is a better option than separate grad program in data science
- Good
- Good idea, but needs to futher clarified. Are we talking about 'fundamentals of DS' (i.e. theory) or 'applications of DS' (i.e. practice). DS is obviously an applied field, but its also very generic, so establishing a designated emphisis would need to be more specific than just DS.
- great
- I am quite positive about this.
- I think this is a better idea than a grad group in data science
- It is a good idea to add DE on Data Science for each of these programs.
- It mostly will enable cross-disciplinary work - very important
- OK to do, but unlikely to have a major impact
- Sure.
- That is also an acceptable option, it would build over existing structures, but funding is an issue.
- This could be a good alternative if GGDS is not established.
- This is an excellent idea!
- This might be a good idea.

Student (N=12)

- Collaboration with industry
- CS, Applied Math, Stat
- CS; Applied Math; Stat
- I think with GGDS we could similarly have a designated emphasis, i.e. this DS-emphasis is equivalent to say that this graduate student is in GGDS
- It would be useful
- It would facilitate enrolling in data science related courses. For example, it is currently almost impossible to secure a spot in ML courses offered by CS deprtment as a student from a different department.
- Might be nice!
- Sounds like a good idea given the current industry focus on data science
- Sure, most PhD degrees cover deeply data science in their field. I think a proper review of published work would indicate whether this emphasis should be awarded to a PhD pursuer.
- To establish such a degree emphasis indeed requires students to take enough credits related to data science, however not many courses in math/stats departments are towards this direction (in particular, how to distinguish this emphasis from the stat degree?). It will be a good discussion to build up a study plan for the target students before creating this new emphasis.
- why not
- Yes but wouldn’t most of the people get the same emphasis in this way?

Opponents of Designated Emphasis on Data Science

Faculty (N=3)

- DE has not worked very well.
- From seeing these designated emphases at other schools, they don't seem to be meaningful
- I am not supportive of this

Student (N=2)

- It is kind of late to do so.
- This seems like a weaker proposal

Appendix F

Raw responses regarding getting more resources and support from the administration.

Faculty Responses (N=7)

- Add a communications specialist
- Focused and interdisciplinary recruitment of faculty members with positions allocated to different department with DS emphasis can generate long term collaborative teams and successes.
- Provide service/consulting in grants with an associated position to coordinate groups and grants
- Summer Classes, Undergrad Research (possibly with grads as mentors) or regular classes in applied DS--get real world data and do analysis.
- The UCD administration must pay more attention to and be involved in our campus-wide data science activities. They need to hire a coordinator of 100% effort who could dedicate to his/her time for our campus-wide data science activities.
- There is huge demand for data scientists and students wishing to enter data science program. That fact alone should lead to increases in support from the university in staff as well as faculty so the course offerings can be increased.
- We need to go for larger overarching grants and/or gifts. The administration can help get those $ here.

Student Responses (N=2)

- Data science shows its value in various applications in industry. Maybe we can enhance the connection with companies outside academia.
- More travel funding and Summer school opportunities.

Appendix G

Raw responses from improving data science research on campus.

Faculty Responses (N=9)

- Better HPC support
- Facilitating shared mentoring of students would be great
- it feels very disconnected, something to bring people together
- It's hard to make suggestions for faculties, but I would like to suggest holding data science symposium for graduate students and encourage them to participate and holding data science competition (with reward).
- more interaction with industry and tech companies to see what are there needs
- Recognize that real interdisciplinary work is hard and under-appreciated. Most of us just try to find nails for our hammers and that has epsilon impact.
- See above--much of modern DS in industry is plugging existing tools into new problems. We should have a class where students learn to do this explicitly.
- Support the CeDAR more.
- We need more hirings in data science. We need people who work on theory of deep learning, reinforcement learning.

Student Responses (N=6)

- DS is about data. I think it would be exciting if we could collaborate more with the industry where they have real world data
- I think it is important to provide research-oriented course. Data science is an very active research area where the frontier of research keeps changing. Therefore, it is necessary to provide courses that can lead interested students to the frontier so they can start interesting research right after the course, rather than spending most of their limited time on finding a path to the frontier.
- Initiate more task-driven research which solve real world problems.
- More faculty that work in data science is great. CS lost a few ML oriented professors.
- special funding for students to join in-person related seminars, conferences, summer schools. invite students/prof/employee from massive ds related area to give a talk or communicate
- We need more interdisciplinary research projects which have influence in other communities.

Appendix H

Raw responses from improving data science education on campus.

Faculty Responses (N=7)

- Events where undergrad/grads can present their research
- Faculty hiring
- Hold annual/monthly data science competitions and encourage students to attend. There is famous data science competition website "Kaggle" (a good model for this).
- more faculty and more courses, smaller course enrollments and more frequent offering of courses
- Organize more discussions/meetings about the graduate level courses related to data science among the related graduate programs and groups
- Start recruiting better at high school or earlier.
- Teach students about things like sprint planning, open-source software development (code reviews, unit testing, etc.), and teach them modern ML tools such as Torch, Pytorch-lightning, DVC, etc.

Student Responses (N=8)

- Coursework should follow emerging trends and be code+project heavy (e.g., use Julia)
- Diversify and modernize graduate level courses
- enhance the communication between grad&undergrad&prof
- Promote rigorous math/stat trainings which indeed help students to stand out in competition.
- Provide more hands-on courses in deep learning stuff
- See previous answer about more faculty in ML.
- We need more courses. For example, for math/stat department their could be a machine learning class for theoretical machine learning, also there could be a ML class in CS department for practical machine learning
- We need more data science and ML related courses offered every quarter.

Appendix I

Raw responses from improving data science activities on campus.

Faculty Responses (N=8)

- Establishing more collaborations with industry
- Getting philanthropy/donations is quite important. This campus needs to determine its niche area that is still related to data science. Otherwise, we cannot compete with Stanford and Berkeley.
- Ideally have a Data Science Center similar (in structure) to the Genome Center
- In addition to education and research, I would like to suggest we should hold a specific career fair for data science (not the university career fair). When I was in GaTech, engineering departments have their own career fairs, which are effective.
- maybe a physical location for people to meet and a quarterly seminar that is not mathematics or statistics or computer science but data science and held at general campus facility rather than a departmental facility.
- Perhaps one nice thing happening in other campuses (UC Berkeley, U Washington) is to allow data science to be part of EVERY course on campus. Data is everywhere and that has created a culture of people wanting to use data science in their specialty. Could we do that here? YES, by adding Data literacy requirements to all majors and all degrees.
- Work with K-12 teachers to identify younger talents.
- You should create larger groups focusing on limited large goals. If you want to get anything meaningful done you should have roughly 12 people working (2 profs + 10 students) on the same big goal. You should look at CMU Delphi as a good example. Increase pay of ML profs by a factor of 2, industry increases it by a factor of 5 and has more data. Make data from outside collaborators more accessible, and help identify problems where ML is actually needed. Too often I just implemented basic things in Python.

Student Responses (N=5)

- Build long term relationships with industries.
- Host some data science related conference in Davis
- It would be better to hold more educational seminars that focus on introducing current active research areas
- More data science related courses should be made available as electives
- We could promote data challenges such as Kaggle. We could organize students (both undergrad and grad) to form groups and participate in them