<u>Sabbatical Applicant</u>

Brendan Patrick Purdy, PhD, MS Data Science

Professor of Mathematics & Data Science Lead

Moorpark College

Semester: Spring 2024

<u>Purpose</u>

The purpose of my sabbatical leave is to investigate and reflect upon the myriad of aspects related to Moorpark College's nascent Data Science program. Upon completion of my sabbatical, I will have specific proposals that upon implementation will benefit our students, Moorpark College, the District, and the wider-community.

<u>Data Science: What and Why</u>

Data Science is a multidisciplinary field that is a combination of mathematics, statistics, programming, and artificial intelligence (AI) coupled with subject matter expertise that uncovers actionable insights from data. The type of AI that is primarily used in Data Science is machine learning, which are AI algorithms trained and tested on data. The large amount of data that is generated by such entities as online retailers, social media, and the government is called "Big Data."

Data Science is considered one of the best jobs in the United States, and based on Moorpark College's own Labor Market Information (LMI) data that we needed to justify the creation of the Data Science Career Technical Education (CTE) Program. Data Science is a

growing field in our geographic region and employers need workers with Data Science skills. Learning Data Science is a great way for those from underrepresented groups to become involved in the STEM fields since it approaches technical ideas from a different manner than is traditionally done due its pragmatic and hands-on nature.

Background: Personal

I have an interdisciplinary academic background (BS in Mathematics and Philosophy; MS in Pure Mathematics; MA in Games, Decisions, & Dynamical Systems; PhD in Mathematical Behavioral Sciences; MS in Data Science). My thesis in Mathematics was on a machine learning model called neural networks and my dissertation applied another machine learning model (decision trees), mathematical logic, and mathematical linguistics to a statistical model used in the cognitive sciences. Additionally, for a decade and half, I have been a Statistical and Data Science consultant for entities in education and industry.

This is my 15th year at Moorpark College as a Mathematics Faculty member, where I have many times taught the classes most germane to Data Science, i.e. Statistics, Discrete Mathematics, Calculus, and Linear Algebra. I have also been deeply involved in the Curriculum for all of these courses.

Recently, I returned to school to earn a Master's of Science in Data Science from the University of Notre Dame. I chose to do this since I realized that I needed Data Science skills to better serve my students and clients. Since my graduation in May 2022, I have continued to reflect on how what I learned from the Notre Dame program can be implemented into Moorpark College's Data Science program.

Background: Moorpark College's Data Science Program

In May 2018, I attended the Two-Year College Data Science Summit sponsored by the National Science Foundation and the American Statistical Association. At this working conference, those in attendance considered three community college student populations:

1. Those seeking employment following an associate's degree

2. Those seeking transfer to four year programs

3. Those seeking certificate programs and college-level courses in Data Science for professional development.

It was decided in consolation with a myriad of stakeholders at Moorpark College, that we would begin with a focus on (3) by beginning a Career Technical Education (CTE) Program in Data Science.

In spite of the lengthy bureaucratic process and the COVID-19 pandemic, under my leadership, the Data Science CTE program at Moorpark College was launched this semester, viz. Fall 2022. Our program is the only Data Science program between Santa Barbara and Santa Monica and is unique since it is an interdisciplinary program which offers emphases in seven disciplines: Business, Computer Science, Computer Systems & Network Engineering, Hospitality, Mathematics, Philosophy, and Political Science.

I should note that I have informed my Dean, my Chair, and the CTE Dean of my plans for this sabbatical.

Rationale & Topics of Research

Moorpark College's Data Science CTE program is unique and offers our students the opportunity to learn about Data Science that can help them whether they want to transfer or to "level-up" in their current career.

Rationale 1: Data Science is a dynamic field, and as such the tools and skills that are used are constantly changing. What was innovative a couple of years ago in Data Science, may have been replaced by different methods. When I was learning Data Science during my Master's Degree program, a number of the packages and tools that I learned have now been updated or even replaced. For example in the Python programming language, I am now using the *scikit-learn* package for Natural Language Processing instead of the *NLTK* package. *The topic of research is to investigate the Data Science tools that are being taught and used in government, industry, and education to make sure that Moorpark College's Data Science Program is current.*

Rationale 2: Since Data Science is a new field both in the workplace and in education, while I am an expert in teaching Mathematics and Statistics, it is important that I learn about the pedagogy of Data Science. There is undoubtedly a lot of overlap between how one teaches Mathematics, Statistics, and Data Science. However, due to Data Science including aspects of computer science as well as subject matter expertise, there are also a number of differences *The topic of research is the pedagogy of Data Science so I can not only teach students the state-of-the-art methods of Data Science, but do so using cutting edge didactics.*

Rationale 3: When we began the process of creating the Data Science program, there were but a handful of bachelor's degrees in Data Science, and none in the area.[1] Now, all of the UCs that offer undergraduate programs have Data Science majors, with the exception of UC Merced, which offers a Data Science emphasis. At least a dozen of the CSUs offer Data Science majors or minors. In particular, both CSUCI and CSUN offer Data Science minors. A number of local private institutions also offer Data Science programs. Here is a partial listing. Caltech, Claremont McKenna College, University of Southern California, Chapman University, University of San Francisco, and Point Loma Nazarene University all have Data Science majors while Loyola Marymount University, Pepperdine University, Pitzer College, Scripps College, and Azusa Pacific all have Data Science minors. Further, while they don't have majors, University of San Diego, California Lutheran University, University of La Verne, and Concordia University all offer Data Science programs. *The topic of research is the curricula at four-year institutions in order to align our curriculum with what is being done at the UCs, CSUs, and local private colleges and universities.*

Rationale 4: While Moorpark College has a Data Science CTE program, the program is not truly institutionalized. We do not have a program that has a prefix in the catalog, i.e. there are no classes called DS M*xx*. All of the courses in the program are in other departments since we do not have our own department. Related to this, we do not have a program plan or a physical location to call our own; we cannot tell a student to go to such and such building and room to visit the Data Science office. Additionally, due to the recent proliferation of Data Science programs at colleges and universities that our students our transferring to, it would behoove

---

[1] While many of the names of the programs that follow are called "Data Science," they often go by other names such as "Data Analytics," "Data Information," or "Data Theory."

Moorpark College to create a Data Science transfer degree. While I am currently working with the CTE Dean through Grant Funding on these issues, it will be optimal to be on a sabbatical to better envision how we can accomplish permanence while giving our students more opportunities for success. *The topic of research is to learn how Moorpark College, our sister colleges in the District, and other community colleges have created new programs (Data Science or not) that are sustainable and have better met their students needs as a result.*

Brendan Purdy's Sabbatical Proposal 2023 - 2024

Schedule/Plan for Sabbatical Work

N.B. "R1" refers to *Rationale 1*, "R2" refers to *Rationale 2*, etc.

     In addition to what is in the schedule below, I also plan to attend a conference or two that are related to Data Science such as the American Statistical Association's Symposium on Data Science and Statistics which occurs annually in May.

<u>January 2024</u> - *Methodologies and Curriculum*

- (R1) Investigate current methodologies used in Data Science

- (R3) Investigate Data Science programs

<u>February 2024</u> - *Pedagogy*

- (R2) Building on the initial (R1) and (R3) research, investigate the pedagogy of Data Science

<u>March 2024</u> - *Pedagogy and Curriculum*

- (R2 & R3) Meet and communicate with people who are involved in other Data Science programs and teach Data Science courses

<u>April 2024</u> - *Institutionalizing*

- (R4) Meet and communicate with people who have created permanent programs at California Community Colleges

<u>May 2024</u> - *Follow-up and Findings*

- (R1, R2, R3, R4) Follow-up on any remaining aspect of the four rationales

- (R1, R2, R3, R4) Summarize my findings

<u>June 2024</u> - *Final Report*

- (R1, R2, R3, R4) Write and submit the Final Report to the President, the Chancellor, and any other interested parties, e.g the Dean of the Mathematics Department or the Dean of CTE.

Value of Sabbatical Leave

Value to Students: This sabbatical will be valuable to students since it will help update and expand the current Data Science program to better serve them.

Value to College: This sabbatical will be valuable to the College since a lot of resources have been put into the Data Science program and this will ensure the success, expansion, and permanence of the program.

Value to District: This sabbatical will be valuable to the District since Moorpark College has the only Data Science program in the District, and for that matter, in the area and as such this is a program that can draw students to the District.

Value to Myself: This sabbatical will allow me to continue my professional development in Mathematics and Data Science while also benefiting our students, College, and District.

Consistency with the Mission of the District

The mission of the district is to provide "students, in its diverse community, with access to comprehensive quality educational opportunities that support student learning and student success." The Data Science program allows students from any background to be successful in the STEM fields and to serve their community as a data scientist whether in industry, government, or education. The Data Science program supports students in their goals to learn Data Science and be a successful data scientist.

Previous Sabbatical

I had a previous sabbatical in Spring 2015 related to Math M15: Introductory Statistics. My topics of research all led to actionable results.

• I investigated the use of technology in the statistics classroom. I have advocated and held professional development on the use of computer software and statistical programming languages in Math M15.

• I investigated the pathway from algebra to introductory statistics. As a result of this, I created Math M08: A Pathway to Statistics, which allowed students to take that class instead of the algebra sequence. With AB 705, we had to sadly retire that course. I then created the Math M15 Bridge and M15 Support courses in light of AB 705, which both serve a similar purpose as Math M08.

• I investigated the roadblocks that students have to get to Math M15 as well as to succeed within the course itself. I found that the proper use of technology and allowing students alternatives to getting into Math M15, i.e. the first two bullet points, led to greater student success.

My previous proposal and final report can be found under 2014-2015 [here](#).

Lastly, my previous Sabbatical is what lead me to first learn about Data Science. Thus, a wonderful unintended consequence of my previous sabbatical is the Data Science CTE program that we now have at Moorpark College. I would not be surprised if such serendipity occurs again with this sabbatical.

Achievable Outcomes

- Implement changes to the current Data Science curriculum to reflect the current best practices in the field

- Create new Data Science courses as necessary

- Work with administration to further institutionalize Data Science at Moorpark College

- Offer Professional Development to the faculty that teach courses that are part of the Data Science CTE Program