Applicant: Nathaniel del Rosario | Website: natdosan.github.io | Email: nadelrosario@ucsd.edu

# Statement of Purpose

"The data you will be working with is vital in detecting down-syndrome as early as possible in pregnancy - great question," my manager answered as I learned more about my project at my most recent internship. As someone who was misdiagnosed with down-syndrome, such an unexpected moment confirmed that my reason to pursue data science had come full circle. My strengths lie in the harmonious intersection of data science, creativity, and their applications to different domains. Consequently, I have demonstrated my ability to thrive in this intersection through multiple internships, extracurricular & research projects, and academic leadership positions. Ultimately, I am applying to Berkeley's MEng EECS Program because the curriculum, resources and research opportunities, and emphasis on leadership align with my desire to become an industry leader in data science applications towards improving healthcare & biotechnology, and I have an established history of steps towards this goal throughout my undergraduate studies.

**Industry Efforts.** Beyond my coursework, I have engaged in multiple internships, research projects, as well as leadership positions in and out of the classroom. During my first internship, I developed a database visualization tool to help researchers identify significant genes which contribute to virus (Ebola, COVID variants, Dengue, etc) transmission. Through this I learned about working in an industry setting contributing directly to ongoing research with individuals from Berkeley, Stanford, and UCSF and more importantly how a data scientist can contribute to the big picture solution in a large research project by considering the perspectives of those who benefit most from my work.

After working with graduates from these schools, I knew I was comfortable collaborating in this environment and considered the value graduate school would bring to my future career. In my most recent internship as a data science and bioinformatics intern, I was involved in developing a test to predict down-syndrome during pregnancy so families have enough time to make a decision to terminate or continue pregnancy. In this industry application of data science to healthcare, I felt the most impact and desire to succeed because my impact will be folds greater when my work eventually culminates in a working aneuploidy test for future mothers across the world, further contributing to my decision to pursue graduate studies.

**Leadership** has also been an integral part of my goals as a student and a data scientist. As an instructional assistant for 2 data science courses serving over 1200 students, I've contributed to my department through helping many new students become data and coding literate by guiding them through my past mistakes. From tangible efforts such as guides on how to manipulate data using Pandas, video demonstrations, grading and beta-testing exam, quiz, and homework questions for the course staff, I've provided a sandbox for students to make mistakes and allow them to grow as the course went on. Overall, I believe leadership is a reflection of my career goal - to manage data science and machine learning projects and ensure they are used for improving our lives. Through the culmination of these diverse experiences in industry and academia, I realized that maturing as a data scientist happens when

acknowledging our skills are most valuable when we collaborate and apply our technical skills to domains like healthcare and realize results only numbers and code could lead us to.

**Future Plans & Fit.** This layering of industry and leadership experience has helped me prepare for graduate school and a career beyond. At UC Berkeley, I would be most excited to pursue studies in deep learning architectures and applications (CS 282, 285), cloud computing & infrastructure (CS 267), and more machine learning applications to healthcare. Having worked in industry-research based environments, I know first hand what working with former Berkeley graduate students in a research environment is like and enrolled in CS188 during a summer internship to apply its topics to my project. Outside of these formal positions, I often found myself looking at Masters and PhD students' portfolios at BAIR, Sky, BEAMMO, and CTML. Following the trend of my industry experience in ML for healthcare/biotech, my future research interests focus on algorithmic bias in healthcare ML as well as making AI powered healthcare tools more accessible and reliable.

Specifically I would like to engage in research similar to Dr. Petersen's "Adaptive Designs for HIV Control" (CTML) and especially Professor Irene Chen's initiative on making machine learning in healthcare more robust and equitable (Chen Lab). I aim to answer questions such as "what kind of biases occur and how do they differ in different supervised learning methods", "why do these biases differ between on ethnic majority/minorities?", and "how can we train future models to account for these potential biases?" To prepare for potential graduate research opportunities at Berkeley, I began guided research under Professor Ilya Zaslavsky in which I designed a regression model and robust accessibility metric to fairly score regions on how equitable public transit infrastructure is. Additionally, for my senior capstone project, I am involved in Professor Zhiting Hu's group centered around LLM reasoning to continue pursuing AI applications to healthcare benchmarks. The resources Berkeley provides for multidisciplinary research problems like these are second to none and provide a perfect setting for my intellectual interests to thrive.

To conclude, a masters in the field is an intermediate goal. Referencing my career goal, I hope to explore as many applications and current problems in data and ML to improve its impact on biotech and other related domains. As demonstrated through my undergraduate experiences in industry, academia, and within the student body, I excel in an environment where I can collaborate and share my knowledge to others. Graduate school at Berkeley would provide such an environment in coursework, faculty, and student body where I can continue to mature and learn. Reflecting on my contributions to virus transmission research and prenatal down-syndrome testing through data science, the imprint I left on my students as an instructional assistant, and my future career goals, I am thrilled to pursue a graduate education to continue growing as a student, data scientist, and leader as I transition to industry.