Statement of Purpose

Prathmesh Savale
(MSc Applied Computing, Fall '21)

Introduction: Motivation to pursue an advanced degree

Watching and reading science fiction while growing up, I was always fascinated by black holes, time travel, and parallel universes. This further developed my interest in quantum physics, general relativity, and the decades-old quest to reconcile them. Which is why I closely followed the ATLAS experiment. I couldn't help but gush over the immense scope of experimental physics and numerical computation in seeking answers to the fundamental questions about our universe. Incidentally, the discovery of Higgs Boson in 2012, and the 2014 Kaggle challenge which involved the use of simulated data from the experiment to predict the presence of the God particle is what got me interested in Data Science.

Data Science helps to draw valuable insights using the petabytes of data generated by intelligent systems to solve real-world problems. I wish to gain a deeper understanding of statistics and programming necessary to generate these insights and help businesses make strategic decisions into their problems. The MSc Applied Computing program at the University of Toronto with Data Science concentration will set the right tone and help me develop these skills.

During my undergraduate study in Computer Engineering, I developed a strong understanding of linear algebra, calculus, and probability theory, but found the exposure to the specific courses in Data Science such as Bayesian methods, optimization, stochastic modeling, very limited. I managed to develop a strong skill set in numeric computation in Octave and software development in C and Python. But felt the lacunae in my knowledge in applied statistics necessary for machine learning (ML) or deep learning (DL). This gap in skill set is what motivates me to pursue an advanced degree in computing with a dedicated focus on statistics.

Experiences: Continuous learning from the industry

In the year 2015, I completed my undergraduate degree from the University of Pune, India. My flair for programming, coupled with my keen interest in problem-solving helped me excel in my studies as I secured a *"First Class with Distinction grade"*. I realized that the best testament for my academic knowledge would be an endorsement from industry experience. With my first job at Mu-Sigma, India's largest pure-play data analytics company, I began my journey into the field of Data Science.

As a Decision Scientist, one of my major projects for a Fortune 3 giant in the consumer electronics space involved reducing their product failure rates. This involved building a classification model on device diagnostics information post repair to predict their probability of failure after going back to the customer. Products with a high probability of failure will be discarded before returning to the customer, thus reducing the failure rate. Since a device failure is a rare event I was looking at highly imbalanced classes and majority of the device diagnostics features showed little to no contribution to variance in the data. This resulted in too many false positives in prediction from the model which if used would be detrimental to the inventory as it meant scrapping devices that are still functional. I came up with a novel approach to building a cascade of classifiers that helped optimize for recall and reduce false positives. Since the cascade used XGBoost, the final ensemble was also able to account for the imbalance in classes.

Another project for the UK's largest retailer entailed building a sales forecasting framework. A data pipeline had to be created to forecast the sales for all the products sold by the retailer across the UK and Ireland regions. This project helped me gain hands-on experience in using time series analysis on line level sales data. I also developed expertise in parallelizing parameter tuning and model building using Python

and PySpark to forecast product sales at different levels of the buying hierarchy. In recognition of my performance, I was awarded the Spot Award and the Impact Award during my tenure at Mu Sigma.

In 2018, I moved to Kiewit Corporation to work with its in-house analytics team. As a Data Science Consultant, I was involved in building text classification models for entity resolution of procurement invoices to track fraudulent vendor claims. Another one of the projects had me build a competitive bid model for profitability to model the propensity of Kiewit winning a construction contract with the optimal bid amount. Working with Kiewit on projects that involved natural language processing piqued my interest in using unstructured data to generate business insights. I started to appreciate the potential of DL as a powerful enhancement to the traditional statistical modeling paradigm. Subsequently, I focused my attention to understand neural networks and some of the popular frameworks used to implement them.

To continue working in a role that would help me develop expertise in DL and enable me to employ my skills as a developer, I moved to work with LTI. Subsequently, I also took the TensorFlow Developer Certification test to deepen my understanding of the framework. Currently, I am working on enhancing the DL capabilities of LTI's proprietary product called Leni. My work encompasses building FastAPI based microservices using TensorFlow and scikit-learn for forecasting and anomaly detection. The APIs are integrated into Leni as part of the TFX airflow pipelines and deployed in production.

Goals and Area of Interest

Every project that I worked on revealed a new facet of Data Science motivating me to gain niche knowledge in the area. With data getting complex by the day, and its generation becoming larger than imaginable, there is a growing demand for analyzing data that impacts the business. My area of interest is in developing automated systems to generate business insights using unstructured data. I chart my career goals as a data scientist with leading institutions. After graduation, similar to my current role, I want to continue developing ML integrated product frameworks leveraging autoML for insight generation on structured and unstructured datasets. Five years down the line, I hope to develop my product suite that will help business owners make data-driven decisions with quicker turnarounds, eliminating the standard boilerplate required for prescriptive analytics.

Apart from work, I enjoy speedcubing and reading books on astronomy, physics, and history. I maintain a *blog* that highlights some of my work in Data Science. I also contribute to a few open-source repositories, specifically in areas of ML and autoML on *GitHub*.

Conclusion: Why this program?

After spending 5 years since my undergraduate degree, working on data science specific projects across the industry, I realized that the gaps in my understanding of applied statistics and mathematics necessary for Data Science can only be filled in an academic setting. Specifically, a program offering a dedicated curriculum in statistics along with computation. Because this skill set is also going to help me in my long term goal of developing an autoML based product suit for prescriptive analytics. The MScAC program at the University of Toronto (UofT) caters perfectly to my requirements since its courses are offered by both the Statistical and Computer Science departments. Moreover, UofT is one of the best universities in the world that has a world-class study and research atmosphere, the best faculty, exceptional peers, and excellent infrastructure, an institute that can facilitate resources to help attain my ambitions.

My efforts during undergraduate study and industry tenure to develop specific skills in Data Science highlight my interest in the field. Moreover, the recognition and awards given by my employers prove that I have the mettle to succeed in the highly competitive MScAC program at the University of Toronto.

Thank You for reviewing my application.