**STATEMENT OF PURPOSE**

It gives me immense pleasure and honor to apply for the **Master of Science program in the** **Data science** track at the University of Southern California. This decision was driven by two important factors. The first thing that drew my attention was the assorted subjects that USC Data Science provides. These subjects not only cover the Computational part comprehensively but will also provide me with an in-depth understanding of the Numerical aspects of Data Science as well which is equally important. The second reason is my strong propensity toward computer programming. This avid interest in Computers is responsible for most of my prominent project works till now and also the reason behind the laurels that I have earned doing them.

My first notable project was during my under-graduation when I had gained a deep understanding of the working of an Atmel 8051 microcontroller. This was a multi-disciplinary project which included students from CS, Mechanical and our Electronics and Telecommunication department and was about creating a GPS guided vehicle with a mechanical arm. This arm was controlled using the 8051 micro-controller and stepper motors. I had used my knowledge of the micro-controller to completely program the movement of the arm. After thorough research, I and my team prepared a group thesis paper in which I explained the technical details like components used, the logic behind the movement of the arm, drawbacks, and future improvements. Our paper was selected as the best research paper of the year by the panel of judges that consisted of all honorable HODs from all the engineering departments of our college. Though I have always opted for computer science electives, the above project with the CS department made realize the knowledge gaps that I have with respect to a CS student. Hence, I decided to hone my skills further by taking advanced Computer Programming courses like C++ and Oracle SQL and passed those courses with an A+ grade from a reputed institute. This was the same time when my first Company Tata Consultancy Services Limited conducted a basic level of training and an entry-level examination before joining that consisted of basics of the Operating system, data structures, computer architectures. These courses and training were quite helpful in bridging the gaps to a great extent.

After I joined TCS as an associate trainee I had the opportunity to learn JAVA, UNIX, and SQL as part of their initial training program and completed these training courses with 5-star gradings. After these training, I was selected to lead a group of four others to develop a website for the company to manage incoming couriers efficiently. During the next 10 days, we presented a technical design model of the website. Since it was time to join our respective departments in the company we handed over the project to our immediate juniors with the login portion successfully implemented. Soon after, I joined the ‘Support and Maintenance’ team of Banking and Financial Services business unit where I made use of my knowledge in above-mentioned programming languages in real-world scenarios. I started with automating simple day to day tasks like checking the quality of files delivered to us based on company standards, which critical files are still left to be delivered etc. The number of files was in hundreds and all these were either checked manually or only came to notice once there were failures because of the missing files. Once this program was deployed into production, it helped to save at least 2-man hours every day.

In this same project, we faced problems like missing data, incorrect data that contains special characters, incorrect file format, incorrect column spacing to name a few. Our traditional way of investigating the cause included steps like manual searching and comparing which only resulted in time-consuming fixes. Because of these problems, I realized the consequences of unmanaged data. So, to streamline the process of investigation, I started exploring the different ways in which we can arrive at such solutions in a more simpler way. After talking to experts who are currently working in this field and a thorough research on the internet I found out that Python is quite useful in these situations. I started to enroll in different online courses. After having complete relevant knowledge about Python, I moved towards python modules for large data handlings like **NumPy** and **Pandas**. I started loading the files to be investigated as Data Frames. Once this was done the process of searching missing values, special characters etc was like a cakewalk. In addition, I used **Matplotlib** to draw inferences like which columns in files were generally affected and the frequency of different errors. This helped the project in reducing the incident tickets from **18000 per month** to **12000 per month**. As compared to the plethora of techniques and theories in Data Science, I had just used the tip of the iceberg, but the outcome was drastic. After this experience, my interest in Computer Science, in general, became more specific. And during the last 2 years, I have enrolled in different online courses, studied from books and have done everything in my capacity to implement these techniques wherever possible. But haphazard choices of courses, lack of guidance and very limited scenarios to implement and practice led to a slow progress. To overcome these obstructions, I decided to devote my further dedication to a full-time Masters program. And when it comes to pursuing MS in this field, the Data Science track of ‘**Master of Science in Computer Science’ offered at USC** is a one-stop solution which will not only provide me with a comprehensive understanding of the subject but will also provide me with ample opportunities to practice and apply my learnings.

Since I already have a taste of the kind of problems that I am facing in my corporate stint, I plan to learn the data **analysis** and **visualization** techniques in absolute depth. Later I plan to move to advanced Data Science topics like the different **Regression Techniques** and other Machine Learning algorithms. And the courses like ‘**Analysis of Algorithms**’ and ‘**Foundations of Artificial Intelligence**’ as mandatory courses will help me achieve the same. Adequate hands on is what I believe to be the best way to learn and hence, once I have gained the required knowledge my first aim is to join the prestigious USC’s ‘**Information Sciences Institute’** and learn as well as contribute towards the different Artificial Intelligence research works particularly in Natural Language Processing. I also plan to work as a **Natural Language Processing Intern** in any of the AI behemoths. I am confident that this will be the exact time to get the actual flavor of what challenges lie in front of me. Once this internship is over I would like to work on these challenges and fill the gaps. And by the end of my post-graduation, I am willing to work on large-scale AI projects like the Google Assistant. I am really planning to understand how the language processing algorithms work and I would like to contribute my learnings in future developments of the same. Only after I have contributed a significant amount of dedication towards this domain I would like to move on to explore other innumerable fields especially in ‘**Predictive Analytics**’ where I can use my understanding and knowledge to simplify and enrich both human lives and experiences.

Apart from my technical contributions mentioned above, I was also part of different extracurricular activities during my under-Graduation and my professional stint. During my under-graduation, I was part different organizing committees where I have participated as well as organized different technical as well as cultural events like **C++ code wars**, **Web application workshops, Financial Markets workshop** etc. During my corporate stint, I was also part of a public speaking forum called the **Toastmasters International**. In this forum, I have also served as the **Vice President Membership** and have hosted club contests from time to time where I was awarded the **Membership Award** by **The Club Growth Director of Toastmasters International** for inducting more than 15 members. All these have helped me to become a better communicator as well as a better leader. And these 2 qualities will help me in both on Campus and employment with any future companies in fields like understanding what is being expected, what needs to be delivered and how efficiently it can be delivered.

My decision to pursue MS is not because of an overnight whim but was developed over a period of time where I have always strived to understand the science behind computation and applied my understanding every now and then to solve real-world difficulties. Because of this appetite for knowledge in this area I have always opted and excelled in all the CS subjects that were offered in our department during under-graduation. Later, working in the corporate world has shaped my idea from an avid learner of Computational Science to a future Data Science enthusiast. My belief in all-round development has also helped me fine tune my interpersonal skills through extra-curricular activities. Having said all these, I strongly desire to pursue my Masters in Computer Science in the Data Science track at the University of Southern California which will not only enrich my knowledge in the field and help me to take up greater challenges in future.