

## Statement of Purpose

---

While at the National Taiwan University, I studied computer science broadly. I ranked among the top 10 percent in almost all computer science courses and especially excelled in theoretical and graduate-level courses. Knowing that some computer science fields require a strong background in mathematics, I chose to pursue a minor in mathematics. Furthermore, I interned at companies and participated in projects in research institutions to learn the practical aspects of computer science. All these experiences convinced me that my interests lie within applicable ideas that have a strong theoretical foundation. I especially enjoyed machine learning theories and their applications in natural language processing and computer vision. I have completed projects, won contests, and published papers in the field of machine learning. I hope to pursue a PhD and continue my studies in related fields of machine learning.

When I was a sophomore, I joined the Language Technology Lab led by Prof. Zhao-Ming Gao, and worked mainly on a Chinese text chunking system. In the system, I implemented grammar-based methods to generate features from text, and then learned a model by a supervised classifier named *yamcha*. Realizing the advantage and power of support vector machines (SVMs) through the experiences, I joined the group that designs the best SVM software, Machine Learning Group led by Prof. Chih-Jen Lin. After joining the group, I have gained a strong grasp in theories and applications of machine learning. I have written and maintained several pieces of widely-used SVM-based learning tools known as *LIBSVM TOOLS*. Additionally, I have studied optimization and nonlinear programming theories. To know how the theories can be applied, I worked on coordinate descent variants to solve SVM and logistic regression models, and conducted experiments using data in various domains to support the theoretical convergence bound for the numerical methods. Moreover, my experiences on different data and models inspired me to build applications on text classification and active learning. My colleagues and I won the *Cloud Computing Contest* held by TrendMicro from about 100 domestic teams by proposing a parallel relevant ranking algorithm implemented in Hadoop for ranking malicious level of websites. I proposed a accurate and efficient active learning method in my co-authored papers published in *IJCNN* and workshop proceedings of *JMLR*.

To gain further insight into the applications in machine learning, I interned at Siemens Corporate Research (SCR) in Princeton, NJ, for six months. My project's objective was to predict when and where a potential failure would happen on a turbine using data collected from multiple sensors. I investigated how methods like Gaussian mixture model, support vector regression, and k-means can be applied to predict data values in a time series. I also partitioned the database appropriately so that the data processing time was 100 times faster than it had been. The most enjoyable part at SCR was talking to the researchers about their research and projects. They frequently gave me ideas on how to improve my project; I was glad to have help them likewise. In my leisure time, I helped a computer vision group at SCR review papers, and discovered how machine learning is widely applied in computer vision.

I would like to introduce a few attributes about me that display my personality and

my preferred style of working. Firstly, I like to try my own hands at solving problems before consulting others. In my opinion, some great original ideas may be lost if I consult others immediately upon encountering a problem. For instance, the winning strategy on the Active Learning Challenge (see my resume) is actually the first thought that entered my mind before surveying papers and books. Secondly, I enjoy teaching, which is a wonderful way to strengthen my knowledge. I led several study groups with my classmates to discuss advanced materials not taught in our courses. Furthermore, I was selected to be a TA in calculus in my senior year. Rare undergrads were given the occupation, even for mathematics majors. Thus, I treasured the opportunity and worked extremely hard to make seemingly tough calculus tractable. While preparing teaching materials and grading homework, problems in my thinking were often revealed and this gave me a chance to improve. My teaching experiences gave me a solid foundation on various subjects, especially mathematics, and trained me to write easy-to-understand slides, documents and reports. Finally, I am a strong implementer already familiar with many software development technologies and tools. I have spent my spare time learning implementation techniques from web-resources and experienced members from institutions and companies.

My ultimate research goal is to build useful theorems and software applied easily and widely. For me, theorems, especially mathematical models and proofs, are intrinsically fun and give us confidence to work successfully on applications; while software enables people to easily utilize the ideas without understanding the theories. I believe it to be one of the fastest ways to make the world advance. That belief motivated me to write pieces of *LIBSVM TOOLS*, to intern at SCR, to participate in contests, and to choose my current research projects. My career goal is to be a professor. Via my many years of teaching experiences, I am sure that I am enthusiastic to pass down knowledge, and excite the next generation of students that will share my passion for this field.

To achieve my research and career goals, I hope to study in an environment full of active learners and creative ideas. The Massachusetts Institute of Technology is famous for solid training in mathematics and engineering and for having abundant research results in theory and application of machine learning. Thus, I hope to join your PhD program to gain further training and to share creative ideas with other researchers.