**Zhiqiang Sun**

Omaha, NE | 402.999.1318 | [sunzhiqiang04@gmail.com](mailto:sunzhiqiang04@gmail.com) | [GitHub](https://github.com/stars/sachenl/lists/final-projects) | [Blog](https://sachenl.github.io/) | [LinkedIn](https://www.linkedin.com/in/zhiqiang-sun-290b7174/?original_referer=)

Data scientist with experience in data acquisition and data modeling, statistical analysis, machine learning, deep learning, and NLP, a background in scientific research, and a Ph.D. in Biophysics. I bring strong skills in investigating and solving complicated problems that need complex logic and mathematical analysis. Native Mandarin speaker, Street Basketball fan.

**TECHNICAL SKILLS**

Python (Pandas, Sklearn, Numpy) ♦ Data Visualization ♦ Statistics ♦ Data Mining ♦ Natural Language Processing ♦ Deep Learning (Keras, TensorFlow) ♦ Spark ♦ Hypothesis Testing ♦ Mysql ♦ Git ♦ AWS ♦OriginLab (scientific) ♦ Matlab ♦Java ♦R

**TECHNICAL PROJECTS**

**Identify skin cancer with machine learning -** [**Github**](https://github.com/sachenl/project5/blob/main/README.md)

*Diagnosed skin cancer pictures with deep learning and identify the skin cancer type*

* Conducted exploratory data analysis and produced visualizations using Matplotlib and Dash plot to identify distributions
* Utilized data augmentation, and random oversampling for class imbalance
* Trained deep learning models (pre-trained networks, convolutional neural networks) to diagnose skin cancer
* Adjusted model architecture and tuned hyperparameters to best-fit models to the data with 75% accuracy

**Customer Churn Analysis -** [**Github**](https://github.com/sachenl/dsc-phase-3-project)

*Analyzed the churn data of a Telecom company and provided suggestions for customer retention*

* Loaded and cleaned the churn data with Pandas
* Deployed Logistic Regression, k-Nearest Neighbors, Decision Trees, Random Forest, and Support Vector Machine to identify the best model
* Compared ensemble performance through confusion matrices and optimized the best model
* Strategically selected features that produce insightful feature-importance for customers to stay with their plan

**House price predicted -** [**Github**](https://github.com/sachenl/dsc-phase-2-project)

*Built a linear model that can predict the house price and evaluated the importance of house features*

* Load, clean, and reshape data while producing clean numerical results
* Build the linear model and figure out the feature’s importance for the house price
* Analyzed the dependence of house price on the features and provided guidelines for homeowners to buy and sell houses

**EMPLOYMENT HISTORY**

Research Instructor, **University of Nebraska Medical Center**, Omaha, NE 07/2013 - Present

* Investigate the interaction between DNA and various proteins such as DNA replication proteins, HIV restrictions, and DNA rescue proteins using an atomic force microscope
* Use Excel and Origin to analyze 1000+ data points to determine if the various combinations of factors produce a novel result
* Alter, build and improve high-speed AFM prototype to provide required use for specific research
* Innovate hardware to increase usefulness for research including growing EBD tip for high-speed AFM with scanning electron microscope
* Collaborate and coordinate with several teams doing related research with different tools
* Create and deliver presentations weekly to the team and annually to University at large.

Postdoc Fellow, **University of Oklahoma**, Norman, Oklahoma 09/ 2010 - 06/2013

* Investigated the DNA packaging by medicine/lipid and condensin proteins with magnetic tweezers
* Wrote new methods with Matlab and Java for image acquisition and data analysis

**EDUCATION**

**Flatiron School, New York, NY**  05/2022

Immersive 600+ hour Data Science program

**Chinese Academy of Science, Beijing, China** 07/2010

PhD: Biophysics, Soft Matter, Physics

**Shandong University, Shandong, China**  07/2004

Bachelor of Science: Physics