

# SHITENG YANG

☎ 585-764-8189 ✉ [yangshiteng007@gmail.com](mailto:yangshiteng007@gmail.com) [in linkedin](#) [github](#)

## Education

### Doctor of Philosophy (Ph.D.) in Statistics

Oklahoma State University

Aug. 2016 – Dec 2021

Stillwater, OK

### Master of Science (M.S.) in Applied Statistics

Rochester Institute of Technology

Aug. 2014 – May 2016

Rochester, NY

### Bachelor of Economics (B.Ec) in Statistics

Hangzhou Dianzi University

Sep. 2009 – May 2013

Hangzhou, Zhejiang, China

## Technical Skills

- **Programming:** Python(eg. pandas, pyodbc, numpy, scipy, matplotlib, re, scikit-learn, flask), R Studio(eg. ggplot2, mlr3, caret), SQL(MySQL & MS SQL Server), Visual Studio Code, Anaconda, Jupyter, Spyder, SAS, Matlab, LaTeX, MS Excel
- **Statistics:** Statistical Theory, Probability Distribution, Hypothesis Test, A/B Testing, Regression Modeling, Data Mining
- **Machine Learning:** Linear Regression, Logistic Regression, Naive Bayes, KNN, SVM, Decision Trees, Random Forest, AdaBoost, Gradient Boost, XGBoost, Neural Networks, K-means, Hierarchical clustering, DBSCAN, PCA, Isolation Forest
- **Certification:** Deep Learning Specialization (Online Courses on Coursera)

## Working Experience

### Amrock, LLC

Data Scientist

Jan 2022 – Current

Detroit, MI

- Utilized Python to implement exploratory data analysis, algorithms optimization and machine learning models development (eg. XGBoost) for a total of 9 modules in an API platform that provides client data auto-validation service, which greatly improved the performance of this platform by reducing the overall false passing rate from 30% to 4%
- Built a data access module enabling the client data auto-validation platform to independently access third-party data sources, resulting in a significant reduction of the data missing rate from 12% to 3%
- Deployed a data extractor service to production to help business team automate the parsing and typing process of a business clause, lowering the error rate from 11% to only 1% and reducing the workload on policy examiners
- Query, mine and analyze large complex data sets using SQL and Python to answer business questions

### Rocket Companies

Data Scientist Intern

May 2021 – Aug 2021

Detroit, MI

- Completed an urgent request to fix an issue of incompatibility with Scikit-learn tree models by utilizing the Python sklearn-json package in just 2 days, which was much earlier than the expected deadline
- Wrangled large, complex data sets to provide insight into opportunities to improve or automate the business processes

### Oklahoma State University

Data Analyst

May 2019 – August 2019

Stillwater, OK

- Supported teaching faculties in department of statistics on a research project that aims to evaluate the overall performance of undergraduate students enrolled in statistical courses through data analysis and statistical modeling
- Utilized R studio to perform exploratory data analysis, such as, data visualization and missing value imputation
- Trained three statistical models (linear, logistic and ordinal regression) to detect the influential features on students' course performance, and proposed suggestions that led to 26% increase in the course pass rate

## Research Experience

### Parameter Estimation with Monte Carlo Sampling

Oklahoma State University, Ph.D. Research

Jun 2018 - Dec 2021

Stillwater, OK

- Developed a parameter estimation algorithm for economic models based on EM (Expectation-Maximization) algorithm and Monte Carlo sampling, which reduced the computation time by 60% and improved the performance by 70%
- PhD Dissertation: Maximum likelihood estimation under efficient importance sampling for non-linear state space models

### Singer Identification with Statistical Modeling

Rochester Institute of Technology, Master Research

Jun 2015 - May 2016

Rochester, NY

- Employed R studio to build a Gaussian mixture model for the detection of singers' identity based on singing voice data, achieving 96% identification accuracy which is 16% higher than the models based on machine learning techniques
- Master Thesis: Statistical Approaches for Signal Processing with Application to Automatic Singer Identification