

Chao Shi

Data Scientist | Machine Learning | Parallel Computing

New York, NY - Email me on Indeed: [indeed.com/r/Chao-Shi/54ecab10a0fcbb4b](https://www.indeed.com/r/Chao-Shi/54ecab10a0fcbb4b)

Data scientist with experience deploying models to drive operating and financial results. Proven track record as project leader / mentor / key contributor delivering business impact.
Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist

NYC Data Science Academy - New York, NY - April 2017 to Present

- Built a real-time recommendation app in Spark ecosystem. Implemented collaborative filtering, content based NLP similarity calculation, and social graph deep dive algorithms. The team of 5 collaborated in a startup-like environment
- Placed Top 1% in Kaggle open competition: Sberbank Russian Housing Market by combining fast-track tree-based prediction, time series analysis and linear regression, industry level hyper-parameter tuning (Luigi), and Agile development and collaboration strategy
- Developed a geospatial decision making framework implemented with R and Shiny, which is designed to be a general and highly reusable tool
- Scraped Twitch.tv channel info and conducted social graph analysis on streaming teams, demonstrating genuine interest and capability in graph related business cases

Geophysicist

Hess Corporation - Houston, TX - October 2011 to December 2016

- Extensive seismic image processing and illumination modeling experience with terabyte data
- Increased SEC book value by \$60MM for Conger field by quantifying illumination artifact
- Drilled 2 discovery oil wells as lead geophysicist in Shenzi North exploration
- Awarded opportunity to mentor early career geophysicist, chosen as a technical recruiter

Software Development Intern

ExxonMobil Upstream Research Company - Houston, TX - May 2010 to August 2010

- Developed methods to automatically integrate conceptual geologic models into faulted structural framework through a rapid PDE based volumetric mapping technique
- Delivered prototype serial and parallel code, algorithms deployed in production code

EDUCATION

PhD in Computational Geodynamics / Mechanical Engineering

Cornell University - Ithaca, NY
2011

BS in Geophysics

Peking (Beijing) University - Beijing, CN
2006

SKILLS

AWS (1 year), Python (2 years), R (2 years), SQL (2 years), Matlab (10+ years), Machine Learning (2 years), Spark (1 year), Hadoop (1 year), MapReduce (1 year), Deep Learning (1 year)

LINKS

<http://github.com/chao-shi-git>

<http://linkedin.com/in/chao-shi>

ADDITIONAL INFORMATION

TECHNICAL SKILLS

Programming: Python (NumPy, pandas, sklearn, Scrappy), R (Shiny, dplyr, caret), SQL, Matlab, C++

Big Data: Apache Spark, Hadoop MapReduce, Hive, Pig, Kafka, AWS, MongoDB

Machine Learning: GLM, Random Forest, Boosting, NLP, Neural Networks, SVM, PCA, Clustering