Xiaoyu Sun
Data Scientist - Analytics Team

New York, NY - Email me on Indeed:

Willing to relocate: Anywhere
Sponsorship required to work in the US

WORK EXPERIENCE

Data Scientist
Analytics Team   -   New York, NY - February 2016 to Present

1. Statistical Modeling
● Credit Score for small business. BA Score TM
Improved model with features from multiple data sources, like user behavior or bank statement data. Produced reports of model validation on performance and benchmarking, calibration testing with banks data.

● Predictive model of business revenue for Paychex based on US census data.

● Vehicle Pricing model for Advantage.
● Predictive model for potential default loans at different stages, like submitted, 3 months after funded.
2. Portfolio Management Optimization
● Restructured deal status following MECE principle.
● Standardized Internal Rate of Return (IRR) calculation methodology.
● Adapted New Financial Reporting Standards, which standardized the measurement of slow payer and default loans.
● Designed new cohort analysis template and visualization features for automated servicing module and online portfolio
management system.
3. Data to Insight
● Conducted Risk Metrics and Analysis for both internal underwriting, business strategy and investors.
4. Data to Action
● Cooperated with underwriting team working on improve the scalability of internal underwriting process by auto
decision, like autoapprove or autodecline.

● Developed Configurable Decision Tree Model Based on Credit Score for small business.

Data Scientist
biz2credit   -   New York, NY - February 2016 to Present

A research proposal on changepoint problems in modeling high frequency financial data.

● Summarize papers about changepoint problems in time series and work on solving the changepoint problem in onestep semiparametric estimation of GARCH model.

Database Project, The USDA Human Nutrition Research Center on Aging - January 2015 to May 2015

Built a database and a website open for lab members that corporate several public resources and can be queried to produce genecentric reports.

● Using front end technologies: html, javascript, bootstrap, Ajax, google charts ;

using back end technologies: Python, cgi ; Database: MySQL under Unix system.

EDUCATION

M.A. in Mathematics
Boston University, Graduate School of Arts & Science   -   Boston, MA
September 2014 to January 2016

B.S. in Mathematics and Statistics
Peking University
September 2010 to June 2014

LINKS

https://github.com/peddlefish

https://www.kaggle.com/peddlefish

https://www.linkedin.com/in/xiaoyu-sun-b80488a2

ADDITIONAL INFORMATION

Areas of Expertise: Data Analysis, Statistical Inference, Predictive Modeling, A/B Testing, Data Mining, Supervised and
Unsupervised Machine Learning, Graph Analytics for Big Data
Math/Computational Skills: Linear Regression, Generalized Regression, Time Series, Probability Theory, Estimation Theory,
Hypothesis Testing, Data Structure, Big Data Modeling and Management Systems, Big Data Integration and Processing
Database: MySQL, MongoDB
Framework & Tools: CGI, clientserver, Hadoop, PIG, Hive, Spark, Git, Latex, Jupyter Notebook, Tableau
Programming: R(ggplot2), Python (Numpy, Scipy, Pandas, scikitlearn), JAVA, SAS, Javascript, Unix Shell Scripting