

BingjingGu

Data Scientist (RISK,anti-FRAUD)

Address

Shanghai
P.R.China

Tel

+86(0)15901802846

Mail

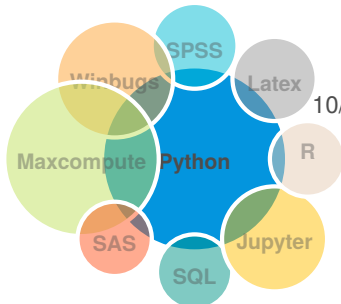
gubingjing@hotmail.com

Web & Git

www.gubingjing.com

github.com/sangaj

Programming



Languages

Chinese ★★★★★
English ★★★★★★

Experience

11/17 - 03/19

Risk Data Scientist

OYO

Design configurable anti-fraud strategies, features, models in terms of business experience, outliers detection, complex network, survival analysis etc; Webscrape open data, construct dictionary database for feature engineering; Design detecting underground industry users models by machine learning algorithm (LR, RF, GBDT, etc); Utilize data tools from alicloud (Maxcompute, DataV, DMS etc) and Hive to deal with etl and visualization, design and construct feature warehouse for real-time and periodic data and provide fundamental statistics, reports and visualization;

11/17 - 03/19

Risk Data Scientist

ZhongAn Insurance

Design anti-fraud strategy model by exploring contact relation, location information, personas etc in terms of complex network, community detection, semantic analysis, density cluster and etc; Web scrape open data and construct relevant dictionary database to support models, features research and verify data, fix price with business partners; Design, construct and analyze credit decision model by machine learning algorithm (logistic regression, RF, GBDT, etc); Create parallel computing to speed up data manipulation and model running;

02/16 - 08/16

Data Scientist Intern

Schiphol Group

Manipulate data from Cloudera Impala running on Apache Hadoop by Spark, Python and R and prepare data for Hackathon and Kaggle competition; Design research area and methods on the bird data from radar systems and search best predictive model in terms of penalized regression, GBM, SVM and neural network to improve the efficiency of bird control in the vicinity of a runway; Utilize H2O to parallel compute all algorithm on Cluster machines by R and Python;

10/11 - 08/14

Data Analyst

GroupM China (WPP Group)

Design, construct and optimize daily, weekly and monthly database of ratings, which further increase update efficiency by 50 % ; Build a drama ratings predictive model by using linear regression and neural network; Program applicable data cleaning module to improve efficiency by C and Excel Macro; Analyze related ratings, media market and advertisement cost data and explore the value for decision making based on channels, markets, targets and advertisers;

04/10 - 10/11

Staff

Agriculture Bank of China

Collect market data, give personalized wealth management advice to clients, promote financial products accordingly by understanding and distinguishing clients needs;

Project

09/15 - 01/16

Statistical consulting for a PHD fellow

Leiden University

Clean infant babbling data from Phonbank, analyze the difference and evolution in languages by longitudinal analysis, provide a Ph.D.fellow with consulting service in her thesis which has been publishing in Cascadilla Press

Education

2014 - 2017

Master's Degree in Statistical Science

Leiden University

2006 - 2010

Bachelor's Degree in Mathematics

Wuhan University

Certifications

Data Science

John Hopkins University / H2O

Coursera Verified Certificates

Practical Machine Learning on H2O; R Programming; Regression Models; Exploratory Data Analysis; Statistical Inference; Developing Data Products; etc;