

GuBingjing

Data Scientist(Risk,anti-Fraud)

Address

Shanghai
P.R.China

Tel

+86(0)15901802846

Mail

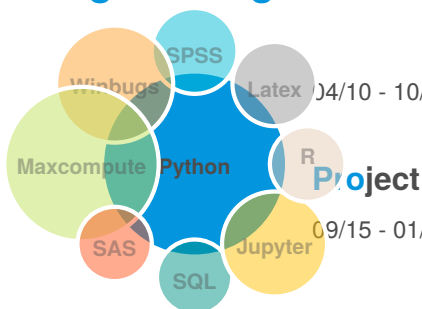
gubingjing@hotmail.com 11/17 - 03/19

Web & Git

www.gubingjing.com

github.com/sangaj

Programming



Languages

Chinese ★★★★★
English ★★★★★

Experience

03/19 - 03/20

Risk Data Scientist

OYO

Design configurable anti-fraud strategies, features, models in terms of business experience, outliers detection, complex network, survival analysis etc to reduce marketing campaign cost;
Web scrape open data, construct dictionary database for feature engineering;
Design machine learning models(LR, RF, GBDT, etc) for detecting underground industry users to improve dynamic pricing strategy and reduce campaign cost;
Utilize data tools from alibaba (Odps, DataV, DMS etc) and Hive to deal with etl and visualization, design feature warehouse for real-time and periodic data and share fundamental statistics, reports and visualization across sectors;

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 dictionary database to support models and features research, validate models, data and fix price with 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;

10/11 - 08/14

Data Analyst

GroupM China (WPP Group)

Program data cleaning module by C and Excel Macro, design and optimize daily, weekly, monthly database of ratings, which further increase efficiency by 50 %;
Build a drama ratings predictive model by using linear regression and neural network;
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;

09/15 - 01/16

Infant Babbling Statistical Consulting

Leiden University

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

02/16 - 08/16

Airport Bird Control Prediction

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;

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;

AML Specialist

ACAMS

ACAMS Verified Certificates

AML Foundations;