

Jian Sun

Data Scientist

Irvine, CA - Email me on Indeed: [indeed.com/r/Jian-Sun/11a16815ac667428](https://www.indeed.com/r/Jian-Sun/11a16815ac667428)

Utilize my expertise in predictive analytics and data science as a Data Scientist to drive business decision.

HIGHLIGHT

- SAS Certified Advanced/Base Programmer for SAS 9; Certified Big Data with Apache Sparks/ Scalable Machine Learning
 - 4 years of experience with R, SAS (BASE/STAT/SQL/MACRO/Hadoop), SQL
 - 2 years of experience with Anaconda Python, Tableau
 - 2 years of experience in relational database: Oracle, SQL Server
 - 2 years of managing/transforming/analyzing big data
 - 3 years of research/consulting experiences in an analytic role
 - 3 years of fraud analysis, credit risk (Chargeback/Credit Underwriting) analysis, forecasting/predictive modeling
 - A Multi-task analyst and a motivated collaborator with quick learning curve and think outside the box
- Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist

Paysafe - Irvine, CA - February 2017 to May 2017

Predict chargeback behavior of different merchant types and verticals with Merchants Categories Code by using

Time Series analysis and Time to Event Survival analysis in R

- Use chargeback predictive model to forecast portfolio chargeback ratios and identify potentially bad merchants to prevent potential chargebacks and losses
- Identify and understand fraud behavior and what types of transaction lead to fraud by creating logistic model
- Perform in-depth data analysis on risk profiles of existing merchant accounts and prior losses to identify new actionable trends by using Tableau
- Prepare and extract historical transaction data from SQL Server. Analyze hundreds of variables to identify emerging risk patterns and create new variables pertaining to these

Data Analyst

FINRA - Rockville, MD - August 2015 to January 2017

Mine and analyze data based on Trader/Broker regulation to detect potential fraud by implementing Latent Semantic Analysis(L.S.A.) in R and Anaconda Python

- Successfully build up User Interface (U.I.) by using Shiny Application to detect text similarity for generic development. The application is accepted by management board and is widely used by department
- Unify item contents using Natural Language Process(NLP) and text mining
- Use machine learning (random forests) algorithm to determine optimal predictors for identifying broker-dealers who are susceptible to complaints
- Build logistic regression under shrinkage methods (ridge regression, lasso), KNN in R and Python language (scikit-learn, pandas, numpy) to perform cheating analysis and circle round suspected candidate

- Identify points of interest by derivatives through nonlinear modeling in R using parametric (logistic model) and non-parametric (smoothing splines) approaches
- Apply Time Series analysis to forecast future exam volumes behavior for CE Online in different vendors
- Create ad-hoc reports by composing query in Oracle SQL and summarize results in pivot table
- Perform data visualization in Tableau to show exam delivery volumes and pricing analysis in different segmentations present in the meeting
- Generate weekly/monthly utilization report including pivot chart, pivot table, slicer, financial analysis and mathematical calculation in Excel to drive business decision
- Research Hadoop platform, spark, Pig, Hive and evaluate potential benefit brings to technology

Data Engineer

Baanyan Software Services, Inc - August 2014 to August 2015

- Composed SQL query in Oracle SQL Navigator to update solution software and improved Utilization Report in Pivot Table
- Coded in Java/Groovy to parse JSON / XML files
- Performed Time Series analysis in R to predict iLink and Netlink customer volumes
- Conduct data visualization with Tableau/Excel

Marketing Data Analyst

- Pulled large volume of data using query tools(SAS/SQL) from insurance database
- Utilized SAS to clean and manipulate large datasets
- Imported Excel format insurance data into MS Access and established Enhanced Entity Relationship (EER) diagram
- Conducted machine learning based predictive model to drive business decision and predict the future market trend
- Used Output Delivery System (ODS) facility to write an analytical report directing SAS output to HTML file which includes statistical tables, analysis summary, and data interpretation
- Analyzed the quality of the output from analytical reports and discussed with business users for solutions to improve data quality
- Quickly learned and mastered the Aviation/Marine/Specialty insurance policies
- Participated in global conference calls to review progress of ongoing quality control of delivered reports

Statistical Research Assistant Intern

Maternal Blood Pressure Analysis at Public Health Department of UC Berkeley - Berkeley, CA - May 2013 to August 2013

- Utilized SAS/Base to clean and transform large datasets, and used PROC SQL to create tables
- Conducted Logistic Regression in SAS to predict the significant factors that affect mortality
- Generated residuals diagnostic to calculate the p-value for testing outliers, produced t-test based on classification variables
- Proceeded ANOVA to evaluate the equality of mean values for mean systolic and diastolic variable across the category
- Initiated SAS Macro to create table that displays the median systolic and diastolic measurements for each quartile of women's body mass index (BMI)
- Composed an analytical report which includes statistical results tables, analysis summary, and data interpretation

RESEARCH/PROJECT

Big Data/Machine Learning Project: Statistical Prediction and Machine Learning analysis for prostate

EDUCATION

B.A. in Statistics & B.S. in Math

University of California, Berkeley - Berkeley, CA
August 2012 to May 2014

Statistics

Santa Monica College (SMC)
February 2010 to June 2012

SKILLS

data visualization (2 years), Excel (2 years), machine learning (2 years), SQL (3 years), visualization (2 years)

ADDITIONAL INFORMATION

SKILLS

Technical: SAS (BASE/STAT/SQL/MACRO/Hadoop), R, SQL, PL/SQL, Python, Java/Groovy, Tableau, Apache Spark, RDBMS(Relational Database Management System), SPSS, MATLAB, VBA, Excel, PowerPoint, Outlook, MS Access, Prezi

Analytic: Predictive Modeling, Scalable Machine Learning, Data Analysis, Data Visualization, NLP(Natural Language Processing), Text Mining, Stochastic Process, Time Series, Game Theory, Financial Analysis, Business Analysis, Operation Research