

Xiao Wang

Senior Data Scientist - Office Depot, FL

- Email me on Indeed: [indeed.com/r/Xiao-Wang/c0e334d9dfeeb88f](https://www.indeed.com/r/Xiao-Wang/c0e334d9dfeeb88f)

- Ph.D. in mathematics; 9+ years overall experience, 5+ years professional working experience in Data Science, Statistical/Mathematical Modeling and Simulation, Algorithms Developing.
- Expertise in problem solving. Excellent in working experience in business environment. Good at gathering modeling needs and converting them into technical solution and data models.
- Deep understanding of Statistical Models and Multivariate Analysis. Familiar with model selection, comparison, validation and testing.
- Experienced in Machine learning techniques: ANOVA, PCA, Forecasting, Time Series Regression, Linear/Nonlinear Regression, Logistics Regression, Clustering, and Tree based models.
- Experienced in working with R and SAS for Statistical Analysis, along with Teradata/SQL Server for Stored Procedures, Functions and Performance Analysis; Hands-on Python.
- Extensive experience in feature collection and selection. Good at data transformation and normalization, variable analysis and visualization.
- Hands-on experience using Hadoop Hive for data management, expertise in dealing with big datasets.
- Experience in Linux Shell Scripting. Familiar with SSH secure shell and remote computing.
- Experienced in parallel computing methods to maximize calculation speed and efficiency, familiar with LAPACK/BLAS. Knowledge in mathematical algorithm realization.
- Great interest in Data Science. Great interest in developing related skill sets and trying new technologies. Fast learner.
- Experience with sensor data collection and analysis from experimental structures and civil infrastructures. Carried out failure model analysis, fissure development prediction and maximum load analysis.
- Teaching and training experience in: College Algebra, Trigonometry, Calculus I, II, III, ODE and Statistics. Patient teacher and good at interpreting materials to students for fast and easy understanding.
- Strong analytical, communication skills that helps in communicating with business customers, development team and management.
- Strong multi-task ability and good time management skills. Energetic and enthusiastic for work. Ability to work under pressure with high efficiency.

Willing to relocate: Anywhere

Authorized to work in the US for any employer

WORK EXPERIENCE

Senior Data Scientist

Office Depot, FL - September 2016 to Present

Responsibilities

- Modeling team production support. Maintain and run models for over 8 Contract, Retail, Print and Email campaigns. Got familiar with the database and models in the process.
- Modeling & scoring automation. In charge of whole automation process. Worked with DBMI, SAS_IT, BI team and automated 2 mature modeling and scoring process. Built guardrail and monitoring reporting system to catch errors.
- Customer acquisition/attrition forecasting: collected customer acquisition and attrition data for forecasting analysis. Used exponential smoothing and then ARIMA on the first stage and collected internal project financial data to study the impact of projects to acquisition and attrition. Time series regression model was used and identified several projects that have high correlation, reported for decision on financial distribution.

- Model review: Reviewed print campaign models, including input feature evaluation, data transformation, oversample/under-sample method, approach of the problem (imbalanced data), sampling and segmenting, model selection process and model validation. Finished detailed report and proposed solution.
- Model enhancement: based on model review, added new data features based on new resources (demographic data), switched the response variable to avoid imbalanced data problem. Also employed more modeling algorithms and added model comparison to achieve best prediction results.
- Imbalanced data modeling: tested for methods like segmenting, change of response variable and balanced random forest method to deal with this problem and lowered misclassification rate largely.
- Implemented R under current system circumstance (Linux and SAS server) for analysis together with SAS for more flexible modeling.
- Data validation, attribute collection for customer demographic and online category.
- Environment: SAS Enterprise Guide, EMiner, R, Teradata, AWS, BI ETL, Python.

Data Scientist

Bestbuy - MN - January 2016 to July 2016

Responsibilities

- Worked on commercial data analysis, including data collection, table treatment, database information, problem analysis, solution proposal and implementation.
- Familiarized with database within weeks. Proposed multiple solutions to deal with sales data analysis and methods to calculate baseline and promotion incremental sales.
- Responsible for data collection, cleansing and ANOVA. Designed technical solution roadmap to deal with noise in sales data.
- Finished multiple linear or nonlinear regression models for testing case and generated very good predicting results in scenario tests. Developed automation process for variable and model selection on over 5,000 core SKU's.
- Worked on Noise Reduction methods - exponential smoothing and Fast Fourier Transformation methods - and made comparison with regression methods. Upgraded old methods to double/triple exponential smoothing and achieved better results.
- Utilized Decision Tree and Support Vector Machine on sales data and carried out clustering analysis on promotional behaviors of different SKU's. Achieved good prediction for different promotion media on various products.

Environment: R, R Parallel, Teradata, SAS, Python.

Data Scientist

Citibank - December 2014 to December 2015

Responsibilities

- Worked with data source team and Client systems to obtain data, identify data issues, data gaps, identified and recommended solutions.
- Worked on data manipulation and analysis & accessed raw marketing data in varied formats with different methods for analyzing and processing.
- Involved in performing extensive Back-End Testing by writing T-SQL queries and stored procedures to extract the data from SQL Database.
- Worked with BI team in data investigation, responsible for interpreting data variables, making instructions and data dictionary.
- Responsible in data pattern recognition and data cleaning. Identify missing, invalid values and outliers, analyze and categorize variables of datasets.
- Checked data for distribution, classification, correlation and VIF etc, worked with BI team for Macro-Economy variable selection to get deep understanding of datasets.

- Carried out Regression Analysis with R/SAS, investigated on the model for problems like goodness of fit, over-fitting, Multicollinearity, residual normality, etc. Established our linear model for forecasting.
 - Worked on principal component analysis and logistic regression analysis on datasets. Involved in variable selection, model optimization, comparing and validation. Responsible for model verification for PPNR requirements.
 - Involved in trend analysis and segmentation analysis based on 12 years of monthly-based market data. Provided graphic analysis report to illustrate market trend and object segment behavior.
 - Preparing model instructions and detailed report for Bureau check, including explaining variables and scenario tests of predictive models.
- Environment: R, T-SQL, Teradata, SQL Server 2012, SSRS, Linux.

Research Assistant, Ph.D. Candidate

UT - Arlington, TX - January 2009 to December 2014

Roles & Responsibilities:

- Experience in simulation code development, data generation, processing and analysis of 3D supersonic flow with MVG. Carried out following stability analysis for understanding complex problem.
- Worked on clustering analysis on 30,000 data sets and profiling of turbulent flow data. Extended research to topology and physics analysis based on analysis results.
- Built predictive model with linear regression analysis. Validated models with model comparison and visualization methods. Compared results against wind tunnel test data and achieved good fitting results.
- Utilized mode recognition and FFT methods to process and filter signal and noise produced in computation results. Carried out time series analysis on time step datasets.
- Carried out ANOVA with R to keep track of variables of turbulent flow and analyze influence of dimension and position on flow parameters.
- Employed Unix Shell Scripting and Parallel Computing for R and Fortran on Texas Advanced Computing Center(TACC) to guarantee efficient and successful computation of our case.
- Familiar with computational numerical algorithms. Experienced in advanced high speed computing and efficiency improvement.
- Responsible for data visualization in both graphical form and motion pictures to analyze topology and support future research.
- Employed methods of interpolation, extrapolation and Fast Fourier Transformation to achieve clear working results which achieved quality similarity with experiment data.

Data Scientist

US Bank - Richfield, MN - August 2013 to May 2014

Responsibilities

- Used and supported database applications and tools for extraction, transformation and analysis of raw data.
- Worked with end users to gain an understanding of information and core data concepts behind their business.
- Analyzed business requirements, system requirements, data mapping requirement specifications, and responsible for documenting functional requirements and supplementary requirements.
- Worked on Data Quality Controls and Business Requirements Document. Attended project release meetings.
- Created data masking mappings to mask the sensitive data between production and test environment.
- Identify source systems, their connectivity, related tables and fields and ensure data suitably for analysis.
- Involved in extensive data validation with several complex SQL queries and involved in Back-End Testing and worked with data quality issues.
- Worked on logistic regression analysis on credit datasets. Built logistic models to support Business department in determining credit issue.
- Developed, managed and validated existing data models including logical and physical models of the data warehouse and source systems utilizing SAS.

- Wrote simple and advanced SQL queries and scripts to create standard reports for senior managers.
- Environment: SQL Server 2008, R, SAS, T-SQL, Windows Server 2008

Credit Analyst

Bank of Communications of China - Taian - June 2003 to June 2005

Credit Project Investor

Responsibilities

- Worked with Tai'an City Construction Survey & Design Institute on the project survey and feasibility study. Worked on investment benefit prediction report and project budget calculation.
- Responsible for communication and negotiation with project related aspects on project loaning, construction budget, design alterations , and unexpected events on the project.
- Worked on project cost analysis. Investigated on construction cost models from the bank and from construction company, compared the difference and made relational data sets.
- Completed program in Fortran and VBA to retrieve data from different construction models and make comparison. Assisted bank and simplified auditing process for similar construction projects.

Environment: Windows XP, Fortran, Excel VBA.

SKILLS

FORTRAN (7 years), SCRIPTING (5 years), UNIX (5 years), DATABASE (3 years), SAS (3 years)

PUBLICATIONS

LES Investigation of the Generation of Momentum Deficit in the Supersonic Wake from a Micro-Ramp

March 2013

Xiao Wang, C Liu, Y Yan: LES Investigation of the Generation of Momentum Deficit in the Supersonic Wake from a Micro-Ramp. JMST. d 10.1007/s12206-013-1164-x

The Vortical Structures in the Rear Separation and Wake Produced by a Supersonic Micro-Ramp.

2014

Xiao Wang, Y Yan, C Liu: The Vortical Structures in the Rear Separation and Wake Produced by a Supersonic Micro-Ramp. Flow, Turbulence and Combustion. doi: 10.1007/s10494-014-9531-y

ADDITIONAL INFORMATION

Technical Skills:

Database Tools: SQL Server, Teradata, Hadoop

Data Modeling Tools: R, SQL, SAS, Matlab

Languages: R, SAS, Linux Scripting, , SQL, Matlab, Fortran, VBA

Packages: SSH, R Markdown, R studio, R parallel, SAS BI service.

Operating System: Windows, LINUX/Unix.

Others: MSOffice (Word, Excel, PowerPoint, Access), AutoCAD, Latex.