

Feng Lu, Ph.D.

Tel: 917-855-0882

vanafeng@gmail.com

Work authorization type: Green Card

Education

New York University

New York, NY

Ph.D., Statistics/Biostatistics, 2007-2013, core GPA: 3.77

- Relevant courses: Time Series Analysis, Data Mining.

Columbia University in the City of New York

New York, NY

Non-degree Graduate, Statistics, 2007-2010, GPA: 3.85

- Relevant courses: Categorical Data Analysis, Generalized Linear Models, Survival Analysis, Stochastic Process, Statistical Modeling, Bayesian Data Analysis, Statistical Inference.

University of California, Davis

Davis, CA

M. S., Biology, 2004-2007, GPA: 4.0 in economics, statistics and mathematics

- Relevant courses: Microeconomics, Mathematical Statistics, Differential Equations, Probability Theory

Peking University

Beijing, China

B.S., Biology, 2000-2004, GPA: 3.75

- Honors: The President's Research Fund, 2003. The National Scholarship, 2002&2003. Excellent student, 2000&2003.

Experience

Hualala Big Data Fintech, Beijing, China

Jan,2021-present

Director of Big Data Administration

Deal with different data source agencies to get large amount of data for small business loans in China, for this startup company in fintech. Direct data screening and preprocessing for modeling small business credit risk.

Citigroup, New York

Feb, 2015-May,2021

Statistician, Fair Lending Officer

Use statistical models to replicate Citibank's lending process in various products like credit cards, commercial loans, mortgages, and use statistical tools to identify banks' regulatory risks (fair lending risk) and give advices to business to reduce such risk.

Explore the application of machine learning techniques (GBM, Random Forest, Neural Networks) in fair lending risk analysis.

Dun and Bradstreet, New Jersey

May. 2014-Feb, 2015

Statistician in Predictive Analytics

- Text mining project (performed in R and Python): performed sentiment analysis on customer satisfaction in survey data using Natural Language Processing tools in R and Python; helped to identify potential future customers for Microsoft Azure service using companies' job posts data by various machine learning methods (Naïve Bayes, SVM, logistic regression etc.)

- Corporate credit scoring (performed in SAS, SAS EM, SAS Hadoop, HIVE): using various machine learning method (neural networks, gradient boosting model, SVM, etc.) and model selection techniques to model and predict the probability of default/ delinquency and create company level credit score.

AIG, New York

Jan. 2013-Dec. 2013

Statistical Modeler in Financial Lines

- Directors and Officers (D&O) Liability Insurance loss prediction modeling (performed in SAS and R): performed research on predicting probability and severity of D&O Insurance claims by using various machine learning methods (gradient boosting model, decision trees) and statistical skills (censored regression, survival regression, etc.); created optimized models and performed model validation and stress testing to improve the risk management/loss forecasting/pricing of D&O insurance.

Division of Biostatistics, New York University

Statistical Consultant and Research Assistant

Sep, 2007-Jan, 2013

- Research on missing data imputation: (performed in Matlab, SAS/SQL, R/S-plus)
Explored major imputation methods (EM algorithm, Monte Carlo method, regression, *etc.*); created Multivariate Adaptive Regression Splines (MARS) regression imputation method and validated its superior performance through intensive Monte Carlo simulation studies.
- The New York University Lung Screening Program: (performed in SAS/SQL, R/S-plus, SPSS)
Identified diseases associated risk factors in high dimensional data through intensive statistical analyses, which include: generalized linear models (logistic regression, Poisson regression *etc.*), principle component analysis (PCA), clustering and classification (decision tree, kNN, *etc.*), linear mixed models, time series models (ARCH/GARCH), different sampling methods, model selection and validation techniques.
- Advising on FDA pre-applications for pharmaceutical companies:
Worked as research assistant on clinical trials. Review FDA pre-applications for pharmaceutical companies and gave advice on the statistical parts.

Publications and Conferences

- Thesis: Extension of the Regression Method for Imputation of Monotone Missing Values with Multivariate Adaptive Regression Splines (MARS), with Applications to Systematic- Missing -At- Random (SMAR) Study Designs.
- CT-Scan Screening for Lung Cancer: Risk Factors for Nodules and Malignancy in a High-Risk Urban Cohort, Publishing.
- H2O.AI World, New York, 2019
- Chinese Finance Association, New York, 2010.
- Member of American Statistical Association, 2009.

Certifications and Skills

- SAS Advanced Programmer 2009
Passed CFA Level II Exam 2011
- Strong proficiency in statistical/quantitative analysis
Strong proficiency in statistical programming: R/S-plus, Matlab, SAS/SQL, SPSS; familiarity with python
Strong proficiency in Microsoft Office Suite Software
Fluent in English and Chinese (native)