# Qualifications Summary

I am highly analytical and versatile professional with over five years of industry experience in developing data products with large datasets. I have strong background in applying statistics, computer science, and machine learning approaches to solve real-world problems.

* Bachelor dual degrees in engineering and computer science, PhD in engineering, master in statistics (expected spring 2017).
* Advanced knowledge of statistical methods and data mining algorithms: logistical regression, SVM, random forest, ANNs, gradient boosting, CART, PCA, Naïve Bayes, clustering.
* Proficient in Python (Pandas, NLTK, GENSIM, Beautiful Soup, Scikit-Learn), R, SQL (Teradata, MS SQL Server, SQLite), Tableau, SAS, UNIX Shell, Cloudera Hadoop, Pig, Hive, Spark and PHP.
* Extensive hands-on experience in developing data products using Python and Tableau for over 200+ internal customers.
* Experienced in building data driven solutions for consumer lending: model monitoring and development support, underwriting, credit policy, valuation and compliance.
* Fast learner with strong self-motivation, hardworking, creative; authored/co-authored 7 publications and 2 patents.

**PROFESSIONAL EXPERIENCE**

# Principal Data Analyst

## Capital One Home Equity, Plano, TX Aug 2016–Present

## 

* Lead a data analyst team for risk, compliance and Lean Six Sigma; build data ecosystem of real-time origination process control.
* Working on master thesis about customer reviews analysis using Python text mining packages (NLTK and GENSIM) and data mining approaches (Scikit-Learn: Naïve Bayes, SVM, Random Forrest).
* Pioneered, designed and automated data process for customer email collection, comprehensive validation and data quality control, which was set as an example and standard for other LOBs by the enterprise team.
* Certified Scrum Product Owner; work with software engineer pod for building the new generation of loan fulfillment UI as SME and business logic provider; work with data engineer pod to build Cassandra API for workflow management; solve any data related problems and monitor data accuracy.

Award: Capital One Circle of Excellence, Q3 2016

# Senior Data Analyst

## Capital One Home Loan Operation, Plano, TX Feb 2014-July 2016

* Supported a full spectrum of operation data needs: underwriting, loan processing, quality assurance, customer experience, compliance and audit (HMDA, HPML, Flood etc.).
* Designed and developed data products and efficiency tools (Python, SQL and Tableau) for 200+ loan processors, underwriters and team leaders; helped to improve underwriting SLA adherence from 73% to 90% in three months; helped to double funded loans per FTE in 6 months.
* Managed and coordinated the complete life cycle of data products – from project initiation, empathic design, porotype development, floor test, feedback collection, result monitoring and agile improvement.

Awards: Financial Service Mark of Distinction, Q1 2016

First Place, Capital One 2015 Data Community All Hands Expo Competition, Oct 2015

# Data Analyst Consultant / Senior Data Analyst

## Capital One Home Loan Marketing & Analysis, Plano, TX Jan 2012–Feb 2014

* Lead data analyst to support Home Loan Marketing & Analysis business and statistician teams for loan origination, pricing, credit policy and valuation.
* Worked with data scientists to design and build the data wrangling process to consolidate credit, origination, and loan performance data from different LOBs and systems into one database for new application model building.
* Debugged, streamlined and redesigned the model monitoring and quality control process for home equity account level valuation (ALV) models.
* Created standardized scripts to build data warehouse for analysis purpose using SAS/SQL in UNIX environment with Teradata as database; Build ad hoc reports for loan fulfillment process monitoring using SQL and Business Objects.

Awards: Home Loan Circle of Excellence, Q1 2014

Financial Service Statistician Team STAR Award

# Research Assistant/Associate

## Iowa State University, Ames, IA Aug 2006- Dec 2011

* Applied statistical learning and data mining methodologies to engineering materials design.
* Modeled the nonlinear composition-processing-property relationships of alloys using recursive partitioning, principal component analysis (PCA), artificial neural network (ANN), and genetic algorithm (GA).
* Performed 3-way ANOVA statistical analysis on nanoindentation hardness data of metal thin film.
* Used statistical software and languages: Matlab, JMP, SAS, R.

## Publications and patents: <https://scholar.google.com/citations?user=mZ0UPNEAAAAJ&hl=en>

# EDUCATION

Master in Statistics: Iowa State University, expected spring 2017

Ph.D. in Engineering: Iowa State University, Ames, IA, fall 2011

Bachelor in Engineering and Computer Science (dual degree): Xi’an Jiaotong University (XJTU), Xi’an, China

## 

# Relevant Courses

* Cloudera hadoop training: PIG, Hive, Sqoop, Spark
* Knowledge discovery and data mining, Machine Learning, Statistical Computing with R, Time Series Analysis, Multivariate Statistical Methods, Statistical Methods, Statistical Methods II – Linear Models, Statistical Methods III – GLM and Bayesian Models, Quality Control and Engineering Statistics, Probability Theory, Statistical Inference
* Fundamental of Database Systems, Data Structures and Algorithms, Fundamental of Operating Architecture, Principles of Computer Network
* Certified Scrum Product Owner