Zeyu Dong

41-08 73rd ST, Woodside, NY 11377

(646)378-9924

zd2162@columbia.edu

Education

- **Columbia University,** Graduate School of Arts and Science, New York, NY — September 2014-February 2016
  MA in Statistics, GPA: 3.8/4.0
- **Beijing Institute of Technology,** College of Math**,** Beijing, China — September 2010-July 2014
  BS in Math, GPA: 3.3/4.0

Internship Experience

Findream, New York, USA — December 2016 – Present

Data Analyst

- Extracted personal financial information from different databases and conducted EDA and data mining methods to find potential credit pattern in user data
- Identified the potential credit risks and fraud trends by establishing logistic regression model in Python with 75.4% accuracy
- Improved LR model in Python through a linear combination of XGboosting, Random Forest (decision tree based), and SVM with 5.3% more accuracy
- Visualized analysis results to senior management team to support data-driven business decision

Arecy, New York, USA — May 2016 – December 2016

Trainee

- Established a real-time stock-analyzer platform with data ingestion layer (Kafka), data storage layer (Cassandra, NoSQL database), data computation layer (Apache Spark)
- Created Zookeeper container in Docker machine to assist Kafka to fetch real-time stock price from Google Finance and Redis container to assist Kafka to filter data to dashboard
- Designed front-end dashboard to visualize real-time stock analysis with node.js, D3.js, jQuery and JavaScript for clients

Mass Mutual Financial Group, Hong Kong — January 2013-June 2013

Management Trainee

- Implemented table partitioning on SQL Server to make it more efficient and better resolution for data warehousing and reporting purposes.
- Provided high quality financial data for back-end developers from different databases with SQL

Academic Research

Sales forecast of Rossmann Company, Columbia University — September 2015-December 2015

Team leader

- Conducted exploratory data analysis (EDA) to summarize the characteristic of some features and perform feature engineering
- Applied ARMA (time-series) model to predict the sales in R and enhanced the performance with Gradient boosting algorithm by 4.1%
- Visualized both prediction results in R to support decision making process

Sentiment analysis in movie reviews, Columbia University — February 2015-May 2015

Team Leader

- Transformed review data into numerical data in Python through TF-IDF/W2V and established several statistical models (Random Forest, XGboosting, CNN) to predict sentiment value
- Improved the XGboosing algorithm performance by 5.8% with K-means clustering method in Python

Mathematical Contest in Modeling, Beijing Institute of Technology — February 2013-May 2013

Team Leader

- Developed a web-scraping program in Python to automatically extract data from several websites and joined different tables by region
- Conducted data imputation by KNN and established Gradient boosting model to predict load

Skills

- R (skilled), Python (skilled), SQL (skilled), javascript (intermediate), Hadoop MapReduce (basic), Apache Spark (intermediate), Pig (basic), Hive (basic)