

Eric (Binqian) Zeng

40 Memorial Hwy. Apt. 05R, New Rochelle, NY, 10801

+1 929-208-7103 • ericzengsearchll@gmail.com

<https://www.linkedin.com/in/binqian-zeng-257903126/> • <https://github.com/bz866>

Introduction

Python and C++ programmer. Two years of experience as a Data Scientist with Master's level education background in machine learning, programming, and quantitative finance. Knowledgeable in statistics, data structure, operating system, and optimization. Developed scalable frameworks to automate the risk prediction and signal detection. Constructed production-level pipelines and test reliabilities. Processed large datasets(500+GB) and implemented data mining on text(corporate filings) and image data(satellite images). Configured clusters to support internal cloud computing. Comfortable with a fast-paced and result-oriented environment.

Professional Experience

- Swiss Reinsurance Company Ltd. (Swiss Re)

Armonk, NY
Jan 2019 – Present

 - Data Scientist
 - * **NLP Based Automation Pipeline for Underwriting Solution**
 - Developed an ML pipeline for underwriting automation; scaled the pipeline to 8 primary insurance companies (Python, MongoDB, Git)
 - Implemented data mining and name entity recognition on insurance submissions; trained a regression model that scores submissions to automate the underwriting (Name Entity Recognition, Random Forest, Spacy, Scikit-learn)
 - Developed an interactive web platform to help the underwriting team monitor the magnitude of risk and check underwritten cases
 - * **Wildfire Risk Prediction & Post-Loss Estimation in North America**
 - Developed machine learning model to predict wildfire risk in North America on a 10 km^2 level and automated the post-loss estimation after catastrophes happened (Python, Generalized Linear Model, Random Forest, CNN)
 - Deployed data mining and feature engineering from datasets of meteorology, land usage, transportation, real estates (Zonal Statistics, Pandas, ArcGIS)
 - Large fire prediction performs at precision 37% and recall 81%; narrowed down the loss estimation error within 10%
 - * **Insurance Submissions Database Management & Computing Cluster Configuration**
 - Configured cloud computing clusters to support internal analysis(Linux, Docker, Bash, JupyterHub)
 - Aggregated metadata description for an internal dataset(500+GB); ETL for data mining(SQL)
 - Implemented regression and clustering to identify insurance market opportunities; backtested hypothesis and correlation (SparkML)
- King Street Capital Management, L.P.

New York, NY
Jul 2018 – Dec 2018

 - Data Science Intern (full-time internship for 6 months)
 - * **Companies KPI Forecasting Model**
 - Forecasted KPIs of companies using an ensemble model; all base and top learners are SVMs with different constraints
 - Handled collinearity in meta-features by matrix factorization; reduced overfitting by ridge regularization and noise injection
 - Direction prediction performs at accuracy 86% (SQL, Python, Scikit-learn, Stacked Ensemble Model)
 - * **Analyzing Alpha in Corporate Filings**
 - Predicted if performance of companies is better or worse than the benchmark based on corporate filings
 - Generated sentence representation with word-embedding and captured textual changes over time
 - Achieved accuracy 82% in binary classification with index as the benchmark(Tf-Idf, RNN, attention mechanism, Tensorflow)

Technical Skills

- **Programming Languages:** Python, C++, Scala, R, Matlab
- **Toolkits, Libraries & Databases:** SQL, Spark, Hadoop, AWS, Git, Numpy, Pandas, Scikit-learn, Tensorflow, Pytorch, Keras, Spacy, NLTK, MySQL, MongoDB

Personal Projects

- Operating System Components
 - Keywords: C/C++, Operating System, Object-oriented Programming
 - Simulated the process scheduling(discrete event simulation) with strategies including Round-robin, SRTF, and preemptive priority
 - Implemented the virtual memory manager with page replacement algorithms including random, clock, LRU, NFU, aging, and working set
 - Implemented the I/O Scheduler with disk scheduling algorithms including SSTF, LOOK, C-LOOK, and FLOOK
- Quantitative Investment & Trading Framework
 - Keywords: Market Impact Models, Portfolio Construction, Optimization, NYSE TAQ Data
 - Backed out permanent impact using the Rate-of-Trading Model and optimized portfolio with T-costs using CVXOPT
- Text Summarization Neural Network for News
 - Keywords: Python, Pytorch, Bidirectional-LSTM, Attention Mechanism, Pointer Network, Semantic Relevance
 - A hybrid of extractive and abstractive approaches text summarization neural network (LSTM, Pointer Network. PyTorch)

Education

- New York University, Courant Institute of Mathematical Sciences

New York, NY
Jan 2019 – Present

 - Mathematics in Finance; part-time graduate program on-going
 - Courseworks: Derivative Securities, Operating System(C/C++), Risk and Portfolio Management with Econometrics, Financial Modeling and Computing, Algorithmic Trading & Quantitative Strategies
- New York University, Courant Institute of Mathematical Sciences

New York, NY
Sep 2016 – May 2018

 - M.S Data Science; GPA: 3.3/4.0;
 - Courseworks: Machine Learning, Natural Language Processing, Deep Learning, Statistical and Mathematical Methods, Big Data, Data Science in Quantitative Finance, Advance Python Programming
- Sun Yat-sen University, School of Engineering

Guangzhou, China
Sep 2012 – Jun 2016

 - B.E Engineering Mechanics (Fluid Dynamics Focus); GPA: 3.7/4.0; Honor: Annual scholarship (three years)