

# RAJESH POTHAMSETTY

DATA SCIENTIST

## CONTACT

✉ [rajesh.747.bits@gmail.com](mailto:rajesh.747.bits@gmail.com)  
☎ (310) 402-7095  
📍 Los Angeles, CA  
in [linkedin.com/in/rajeshpothamsetty](https://www.linkedin.com/in/rajeshpothamsetty)  
🐙 [github.com/rpothamLearner](https://github.com/rpothamLearner)

## EDUCATION

2019

Georgia Institute of Technology, Atlanta

Master of Computer Science – Machine Learning Specialization

- Algorithms
- Big Data Application(Spark/ML/Python)
- Software Architecture & Design

2016

University of California, Los Angeles

Master of Mathematical Finance

- Computational Methods
- Financial Data Analytics

2012

Birla Institute of Technology & Science, India

Bachelor of Science in Engineering

- Robotics

## TECHNICAL SKILLS

- Machine Learning
- Time Series Analysis
- Natural Language Processing
- Statistics and Probability
- Data Visualization

## PROGRAMMING SKILLS

- Object Oriented Programming
- Programming LanguageS:
  - Python Stack (Numpy, Pandas, scikit-Learn, Keras)
  - R(dygraphs, ggplot, dplyr)
  - PySpark
- Data Management:
  - MySQL, Apache Spark, Apache-Drill, AWS (Basic)

## EXPERIENCE

March, 2017 - Present

**Financial Engineer / Data Scientist** | Intercontinental Exchange, Inc

- Developed an ensemble model to classify mortgages into prepay, default and payment classes. Built an ETL pipeline to process 20 million records from agencies. Extracted additional features such as economic and borrower demographic variables with importance from FRED, Housing Dept etc. Improved accuracy by 25% in identifying default loans.
- Implemented an on premise server health monitor dashboard which provided intuitive and dynamic visualization of critical variables like (load, memory leakage, heap, thread count) for 500+ servers using markdown, html rendering and dplyr. Processed ~50GB of daily data.
- Built multiple OOP modules to improve the efficiency of ETL process of agency data (~10m records; Fannie, Freddie Mac). Automated data feed quality verification process of daily tolerances, price, agency ratings and portfolio analytics.

April 2018 – July 2018

**Freelance Data Scientist** | Method Data Science, LLC

- Developed a content-based filtering business recommendation system to identify similar users and similar business based on six million Yelp user reviews using NLP modules in PySpark MLlib and Apache-Drill Parquet files for querying. Pre-processed reviews via Regex Tokenizer, Stop Words etc., recommendation via TF-IDF and Word2Vec model with cosine similarity.

April 2016 – Nov 2016

**Quantitative Researcher** | Logica Capital Advisors, LLC

- Created generic modules for data processing, ranking and portfolio generation to accommodate custom risk factors, improved turnaround time by 200%.
- Developed a framework in R for testing risk factors such as price ratios, RSI and price proxy for value and momentum.
- Performed Principle Component Analysis to identify 7 most significant factors out of 19 that explain the variance in returns and applied KNN clustering on 1000 stocks to form co-integrated portfolios.
- Implemented a pair's trading back-testing framework that uses co-integration heat-map of stocks, along with the implementation of Kalman filter and OLS regressions for dynamic hedge ratio of pairs.

## ACADEMIC PROJECTS

- Context based weighting for the prediction of mortality in the ICU using Long Short Term Memory Network and a 5 layered Deep NN with an Adam optimizer and 20% dropout with an AUC ~90% and 98.5% respectively.
- Applied Finance Project (The HOD Group, Luxembourg): Optimized the existing multi-asset allocation system by overlaying dual momentum (VIX modulated) and market timing signals such as Absorption Ratio, Variance Risk Premium. Improved the portfolio performance and drawdown characteristics.