RAHUL, PH.D.

EDUCATION

2012 — **Department of Applied Mathematics, University of Waterloo**
Doctor of Philosophy — Waterloo, Canada

2011 — **Center of Teaching Excellence, University of Waterloo**
Certificate in University Teaching — Waterloo, Canada

2008 — **School of Computational and Integrative Sciences, JNU**
Master of Technology — Delhi, India

2005 — **Department of Electronic Sciences, University of Delhi**
Master of Science — Delhi, India

2003 — **Sri. Aurobindo College, University of Delhi**
Bachelor of Science — Delhi, India

PROFESSIONAL EXPERIENCE

Aug, 2018 | present — **Manager Data Science**
Trafigura Global Services — Mumbai, India

Sep, 2015 | Aug, 2018 — **Data Science Associate**
Blackrock Services India Pvt.Ltd — Gurgaon, India

Jan, 2015 | Aug, 2015 — **Data Science Research Scientist**
Data Science Practice, Impetus Info-tech Pvt. Ltd. — Noida, India

RESEARCH AND TEACHING EXPERIENCE

2014 — **Postdoctoral Research Fellow**
McGill University — Montreal, Canada

Sep, 2012 | Dec, 2013 — **Sessional Lecturer**
Department of Applied Mathematics, University of Waterloo — Waterloo, Canada

Sep, 2008 | Sep, 2012 — **Graduate Teaching Assistant**
Department of Applied Mathematics, University of Waterloo — Waterloo, Canada

CONTACT INFO

+91 8928190300
katyayan.rahul@gmail.com
Full CV
https://github.com/r2rahul
http://r2rahul.github.io/
in.linkedin.com/in/r2rahul/

Trafigura Highlights

Collaborated with Accounting team to build a time-series forecasting model for the price curves

Leading the effort to scope opportunities with trading desks

BlackRock Highlights

Lead the proficient team, which delivered high impact projects like country risk scorer

Streamlined the hiring process using tools like HackerRank and co-ordinated to grow the team from 1 to 11

Impetus Highlights

Lead team to deliver model validation of a catastrophe risk model, which increased efficiency by 33% while improving numerical accuracy on the short timeline

Coached new team members on the principles and art of building production-grade models

SELECTED PROJECTS

Trafigura

Developed a Data Product to Predict Co-Integrated Time Series for Cross-Hedging Future Instruments

Tools: R, Rmarkdown, Docker

- Developed methodology using DTW distance, and cointegration test to identify related instruments
- The model deployed using Docker and used by the derivatives and LNG trading desks

BlackRock

Developed a Methodology to Identify Cohorts of Stock for the Thematic Funds

Tools: Igraph, Spacy, Google Cloud Platform (GCP)

- Used text mining to identify similar stocks from the initial seed stock descriptions
- Constructed a supply-chain and subsidiary company network graph to predict high impact stocks

BlackRock

Designed a Workflow to Identify Industry Categories from Cargo Descriptions

Tools: Python, TextBlob, GenSim, NLTK

- Developed custom word embeddings using vector space model to map descriptions to SIC industry category
- The module was utilized to construct Trading signal with IC .01

Impetus

Built a Real-time Recommender System using Singular Value Decomposition (SVD)

Tools: R, Shiny, Github Source code

- The capability was utilized by the client to access the benefits of the recommendation system
- Used Docker to Deploy the Model in Production

McGill University

Predicted Key Regulators of the Warburg Effect in Cancer Cells using Multi-Scale Systems Modeling

Tools: R, MATLAB, Python, High Performance Computing (HPC), Amazon Web Services (AWS)

- Applied supervised and unsupervised methods to build a knowledge graph of Gene Protein Reaction Association
- Identified important pathways using the aggregated data and model formulated as Convex Optimization objective function

University of Waterloo

Built a Co-Expression Graph from Protein Expression Data and Predicted Functional Hubs of the Network

Tools: R, Igraph, Circos, Bitbucket Source Code

- Adapted Sparse Partial Correlation Estimation (SPACE) to build co-expression network
- Utilized Degree Centrality to Predict Hub Proteins

University of Waterloo

Designed Algorithm for Hyper-Parameter Search through Optimization for the System of Non-Linear Equations

Tools: MATLAB, SUNDIALS ODE Suite, Parallel Computing using Message Passing Interface (MPI)

- Built Optimization Pipeline using Simulate Annealing to find Hyper-Parameter of the Kinetic Model
- Identified key drivers of type-2 diabetes Insulin regulation using the trained model

Techinical Skills

Imperative Programming: Python, MATLAB, R. Also ability with: C, JavaScript, Shell Scripting, Julia

Functional Programming: F#

Specialized Tools: Dask, Pystan-Bayesian Modeling, High Performance and Distributed Computing, OpenMP and MPI, pySpark, Keras and TensorFlow, Prophet, *R data science stack*, *Python data science stack*

Data Pipeline Tools: Elasticsearch, openTSDB, SQL Databases

Production and Reproduciblity Tools: Docker, R Packrat, Python Virtual Environment, Make, Sumatra

Computational Essay Tools: Markdown, Jupyter, R Knitr

Cloud Platform: GCP, AWS

Visualization Tools: Circos, Inkscape, Xfig, Bokeh Dashboard, R Shiny, Plotly Dash

Version Control: Git, Git-lfs, SVN

Educational Tools: Clickers, OpenBoard

R and Python Packages

R package expdata: Developed a light weight exploratory data analysis package using data.table in R

Package Code: https://github.com/r2rahul/expdata

SocialSent Custom Sentiment Lexicon Generator: Upgraded Python Package to work with Python 3 and Keras Functional API

Package Code: https://github.com/r2rahul/socialsent