Rahul, Ph.D.

BlackRock, Gurgaon, India, 122002

+91 9599109605 • katyayan.rahul@gmail.com • http://r2rahul.github.io/
in.linkedin.com/in/r2rahul/ • https://github.com/r2rahul

Education

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

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

Master of Technology **2008**
School of Computational and Integrative Sciences, Jawaharlal Nehru University (JNU), India

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

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

Experience

Industry

Data Science Associate **Sep, 2015 - present**
Blackrock Services India Pvt.Ltd, Gurgaon, India
Description: I am lead data scientist of Gurgaon Active Equities Data Science team. My role includes:

- Maintaining the highest level of work ethics in the team members, well-defined goals, and standard of work for the smooth functioning of the team. Hands-on participation in projects with the team members
- Partner with investors to understand the problem and design project work-flow to decipher insights from data
- Grow team with best skills. Designed an interview process using live coding exercise through HackerRank
- Previously, member of Advisory team. I worked as a lead data science advisor on multiple client projects

Research Scientist **Jan, 2015 - Aug, 2015**
Data Science Practice, Impetus Info-tech Pvt. Ltd., Noida, India
Role Description: As a data science consultant, I was responsible for:

- Conceptualize the client data science problem and coordinate with all stakeholders
- Delegate and prioritize project tasks among team members and ensuring completion of assignments are timely
- Working hands-on with the team to execute projects

Research

Postdoctoral Research Fellow (Research Scientist) **Feb, 2014 - Dec, 2014**
Department of Biochemistry, McGill University, Montreal, Canada
Description: My responsibilities included the supervision of the research team and execution of entire life cycle of the research project. The role involved:

- Maintain data quality, develop and interpret model, and communicate the results to all the stakeholders
- Train lab members, create an efficient work-flow for the faster and timely execution of the projects

Sessional Lecturer **Sep, 2012 - Dec, 2013**
Department of Applied Mathematics, University of Waterloo, Waterloo, Canada
Description: I was responsible for developing lecture notes, assignments, holding office hours, supervising graduate teaching assistants, and rigorously following course time-lines.

Graduate Teaching Assistant **2008 - 2012**
Department of Applied Mathematics, University of Waterloo, Waterloo, Canada
Description: The role involved punctuality in grading student assignments and delivering a problem-solving session to students.

Research Intern **May, 2004 - Aug, 2004**
Indian Meteorological Department, Delhi, India
Description: I developed a linear regression model to predict humidity using temperature and pressure as a feature. The data was collected using carbon hygristors (sensors) from upper air atmosphere. The model was taken to production in the upper air atmosphere laboratory to use model predictions for the further calibration of carbon hygristors.

Technical skills

Programming Languages: Proficient in: Python, MATLAB, R, LaTeX, Octave
Also ability with: C, JavaScript, Fortran 77, Shell Scripting, Julia, Python Dask, pySpark
Cloud Platform: Google Cloud Platform (GCP)
Databases and Literate Programming Tools: SQL, Solr, R Knitr, IPython, Markdown, lintr, Make
Visualization Tools: ggplot2, ggvis, leafletR, Circos, Inkscape, Xfig, dimple.js, Bokeh, R Shiny
Version Control and Project Management Tools: Git, SVN, SourceTree, Sumatra, Jira, Confluence
Educational Tools: *Clickers* an audience response system, *Open-Sankore* an interactive white-board software

Selected Projects

Industry

Network Analysis to Predict Effect due to Cross-talk between Industry Sectors
Tools: Igraph, Spacy, Google Cloud Platform (GCP) *BlackRock*
Description: A network analysis of the Health Care supply chain network to predict critical control points of the network.

Natural Language Processing of financial text to identify financial constraints
Tools: GenSim, Spacy, Bokeh DashBoard, Google Cloud Platform (GCP) *BlackRock*
Description: A custom word embedding was developed to predict financial constraints.

Identifying Industry Categories from Cargo Descriptions using Text Mining Technique
Tools: Python, TextBlob, GenSim, NLTK *BlackRock*
Description: The objective was to determine the industry category like auto, steel, and other from the short descriptions of products. I used Latent Semantic Indexing (LSI) and cosine similarity to predict industry category.

Finding Loan Cohorts for the Efficient Downstream Analysis of Cash Flow
Tools: Python, R, PySpark *BlackRock*
Description: The goal was to create cohorts of the loan, which share similar characteristics. I utilized K-Means Clustering and Principal Component Analysis (PCA) to create loan cohorts.

Real time Recommender System using Singular Value Decomposition (SVD)
Tools: R, Shiny App, (1) Source Code: https://github.com/r2rahul/fastSVDrevisions *Impetus*
Description: The aim was to build a real-time recommendation system based on the web browsing history. Next, I implemented the Brand, M (1) algorithm to construct a recommendation system and R shiny dashboard to showcase the recommendation.

Research

R package *expdata*
Tools: R, R data.table, R Shiny, Source Code: https://github.com/r2rahul/expdata *Independent Project*
Description: A light weight exploratory data analysis tool using data.table in R.

Mechanistic Systems Modeling to Predict Key Metabolic Drivers in the Cancer Cells
Tools: R, MATLAB, Python, High Performance Computing (HPC), Amazon Web Services (AWS) *McGill University*
Description: The goal was to build a model to study the combined effects of gene expression and metabolic changes in the cancer cells. We used PCA, Hierarchical Clustering, SVM, and NLP to aggregate the data for the model. Next, we formulated model as a convex optimization problem to predict key drivers.

Building a Co-Expression Graph of Proteins and Identification of Functional Hub Proteins
Tools: R, Igraph, Circos. Source Code: https://bitbucket.org/r2rahul/sparsecorrelation *University of Waterloo*
Description: The aim was to build a co-expression network using sparse data and identify hub proteins. We used Sparse Partial Correlation Estimation (SPACE) method to build the network. Finally, we identified the hub proteins using degree centrality.

Hyper-Parameter Search through Optimization for the System of Non-Linear Equations
Tools: MATLAB, SUNDIALS ODE Suite, Parallel Computing using Message Passing Interface (MPI) *University of Waterloo*
Description: The problem belonged to a broader class of subject called inverse problem also a sub-problem in machine learning algorithms. The objective was to identify a set of parameters for the system of non-linear equations describing insulin regulation through metabolism. We used global sensitivity analysis, K-Means algorithm, Huber cost function, and simulated annealing to identify the model parameters.

Additional Information

Full CV: http://tinyurl.com/nddsnfe
Teaching/Mentoring Material Samples: http://tinyurl.com/ppxobvd