

Rahul, Ph.D.

BlackRock, Gurgaon, India, 122002

+91 9599109605 • katyayan.rahul@gmail.com • <http://r2rahul.github.io/>
[in in.linkedin.com/in/r2rahul/](https://in.linkedin.com/in/r2rahul/) • <https://github.com/r2rahul>

Education

Doctor of Philosophy <i>Department of Applied Mathematics, University of Waterloo, Canada</i>	2012
Certificate in University Teaching <i>Center of Teaching Excellence, University of Waterloo, Canada</i>	2011
Master of Technology <i>School of Computational and Integrative Sciences, Jawaharlal Nehru University (JNU), India</i>	2008
Master of Science <i>Department of Electronic Sciences, University of Delhi South Campus, India</i>	2005
Bachelor of Science <i>Electronic Sciences, Sri. Aurobindo College, University of Delhi, India</i>	2003
Selected Massive Open Online Course (MOOC) Certifications and Workshops.....	
Scalable Machine Learning (Apache Spark) <i>edX</i>	August, 2015
Participated in High Performance Computing Workshop Series <i>McGill University</i>	2014

Experience

Industry.....	
Data Science Associate <i>Blackrock Services India Pvt.Ltd, Gurgaon, India</i> Description: I synthesize and mentor team members to decipher insights from data. I am responsible for identifying a data set, conceptualizing the problem, coaching junior team members, and hands-on execution of a project.	Sep, 2015 - present
Research Scientist <i>Data Science Practice, Impetus Info-tech Pvt. Ltd., Noida, India</i> Role Description: As a data scientist consultant, I conceptualized the client problem, coordinated between different teams, and developed production level models for the business problems.	Jan, 2015 - Aug, 2015
Research.....	
Postdoctoral Research Fellow (Research Scientist) <i>Department of Biochemistry, McGill University, Montreal, Canada</i> Description: My role involved execution of entire life cycle of model development to understand the complex interplay between metabolism and gene expression in breast cancer. The role involved maintaining strict quality for data collection, develop a model, interpret the model result, train lab members, and communicate the results to all the stakeholders.	Feb, 2014 - Dec, 2014
Sessional Lecturer <i>Department of Applied Mathematics, University of Waterloo, Waterloo, Canada</i> Description: I was responsible for developing lecture notes, assignments, holding office hours, supervising graduate teaching assistants, and rigorously following course time-lines.	Sep, 2012 - Dec, 2013
Graduate Teaching Assistant <i>Department of Applied Mathematics, University of Waterloo, Waterloo, Canada</i> Description: The role involved punctuality in grading student assignments and delivering a problem-solving session to students.	2008 - 2012
Research Intern <i>Indian Meteorological Department, Delhi, India</i> Description: I developed a linear regression model to predict humidity using temperature and pressure as a feature. The data was collected using carbon hygriators (sensors) from upper air atmosphere. The model was taken to production in the upper air atmosphere laboratory to use model predictions for the further experimentation.	May, 2004 - Aug, 2004

Technical skills

Programming Languages: Proficient in: Python, MATLAB, R, \LaTeX , Octave

Also ability with: C, JavaScript, Fortran 77, Shell Scripting, Julia, pySpark, MapReduce Programming Paradigm

Databases and Literate Programming Tools: SQL, Solr, R Shiny, R Knitr, IPython, Markdown, linter, Make

Visualization Tools: ggplot2, ggvis, leafletR, Circos, Inkscape, Xfig, d3.js, Bokeh

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.....

Identifying Industry Category from Good Description 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 Residential Mortgage 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/r2raahul/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.

Modeling Risk due to Natural Catastrophes for the Insurance Premium Estimation

Tools: Python, Cython

Impetus

Description: The objective was to validate the risk model built on the MapReduce programming paradigm. I improved the model using high precision numeric algorithms, converting codes into Cython, and removing redundancy in the algorithm.

Research.....

Mechanistic Systems Modeling to Predict Key Metabolic Drivers in the Cancer Cells

Tools: R, R Bioconductor, MATLAB, Python, High Performance Computing (HPC)

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/r2raahul/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 train the model.

Spatio-temporal Monte-Carlo Simulation to Study Oscillatory Patterns of Min System of Proteins in E.Coli

Tools: MCELL, MATLAB, Python, Parallel Computing using OpenMP and MPI

JNU

Description: The objective was to study the distribution of Min Proteins in E. Coli during cell division. We used Monte-Carlo Simulation and parallel computing using MPI to develop the model.


Predicting CpG island in a Genome Sequence using Hidden Markov Model (HMM)


Tools: R, Bioconductor, HMMER

JNU

Description: The objective was to predict CpG island in a genome sequence. We used the Viterbi algorithm to learn parameters of HMM model.

Additional Information

 Academic CV: <http://tinyurl.com/nddsnfe>

 Teaching Samples: <http://tinyurl.com/ppxobvd>