Rahul Singh

7011832378 ◆ 20-7-83 ◆10/110, Indira Nagar, Lucknow (U.P.) India ◆ rsingh2083@gmail.com

Education

JAYPEE UNIVERSITY OF INFORMATIONS TECHNOLOGY – H.P., India
M.Tech. Computational Biology (2010-12) **89%**
Majors in Machine Learning, statistics, mathematical modelling

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY – Noida, India
B.Tech. in Biotechnology, 2007 **70%**

Employment History

Data Scientist, MindTree, Bangalore, Aug. 2014 – Present.

Awards
Outstanding Performer of the Year (Mindtree): Year 2014 - 2015
Outstanding Performer of the Year (Mindtree): Year 2015 - 2016
Mindtree Pillar Award (2016): Recognizes the top 5% talent of the organization.

- ◆ Built three successful products in a span of two years.
- ◆ Delivered a streak of successful POC's which were shortlisted to be developed as product. Built a team of data engineers & data scientist for our platforms solutions group.
- ◆ Leading a team of 5 on Machine learning, Natural Language Processing and Visualization.
- ◆ Built entire interactive D3js based analytics app for Airline & Hospitality. Wrote various web scrapers to gather data from reviews websites & social channels.
- ◆ Handling interaction with clients, marketing and sales team & promotions on social media. Eliminated need to hire new team members by performing dual roles in marketing and sales, with estimated $40K annual savings.
- ◆ **Deep learning** for text analysis. Complex datasets analysis and simulations of complex systems.
- ◆ **Industries handled : Airlines, Retail, Insurance**
- ◆ Involved actively in guiding all aspects of data cleaning, algorithm design and visualization.
- ◆ Intelligent ticket resolution system to automatically resolve most common user queries, which saves on cost of customer services as well as brings down average ticket resolution time down from 40 minutes to ~5 minutes.

Scientific Support Executive (Discovery Info.), Piramal LifeSciences, **[Operations Closed]**
Mumbai, Mar. 2013 - July 2014.

- ◆ Understanding & assisting a team of senior scientists from pharma, chemistry, mathematics with my machine learning skills to make the most out of the data.
- ◆ Data Analysis on R/MATLAB/Python/Octave. Data volume handled : 80 Million, ~ 1.4TB.
- ◆ Writing interactive statistical web applications in shiny-R for scientists to analyze their data graphically. Mathematical modelling of Metabolic Network Pathways as well as their optimization. Writing programs in perl, python, R for automation & analysis as per the requirement of scientists.
- ◆ Mathematical model for transmission, pattern formation, evolution of complex systems.
- ◆ Extensively worked on **Unsupervised Learning methods**.

- Also handled system administration, software support on Gold, MOE, Schrodinger, Gromacs etc

Statistical Data Analyst (Team Lead), Allusion, **[Startup, Operations Closed]**
Dec. 2011 - Feb. 2013.

- Data analysis on R, Tableau & python. Scientific data analysis & report generation.
- Twitter sentiment analysis: Created python & libSVM based classifier for customer satisfaction and brand management analysis.
- Language Model: Statistical model for language corpus using maximum entropy models.
- Neural Networks implementation for image analysis project.
- Context based grammar correction in documents.

I.T. Engineer, CMC/TCS, **[Left to pursue full time Mtech.]**
Noida (U.P.), Dec. 2008- Apr. 2010.

- **Image analysis**, text analysis. Time Series analysis.
- Data wrangling and sequence preprocessing & comparison for data analysis application.

Kernel Developer & System Admin, Dvico Tech, **[Startup, Operations closed]**
Bangalore, Nov. 2007- Jul. 2008.

- Perl/Shell/Python scripting for cross compilation. Kernel development based on LFS.
- Software defect analysis done along with Korean team members.

Data Analyst, GVG Bytes, **[ShutDown]**
New Delhi, Apr. 2007- Nov. 2007.

- Data Analysis on R/Octave. Installation, configuration & support of statistical tools on Linux Fedora/RHEL. Perl/shell/python scripting.

Freelance Work

- **Machine learning assistance to THSTI (Government of India, Faridabad):** Machine Learning assistance on diabetes related research work, involving temporal patient's records.

- **Interactive Visualization tool for Clips India (NGO, Pune):** Developed a cloud based interactive data visualization tool for NGO's.

- **Healthcare Analytics for Jaypee Hospital (Noida):** Analyzed their in-house patient's records to predict hospital readmissions.

Key Technical Skills

COMPUTER LANGUAGES:	PYTHON, Shell, OCTAVE,R, awk/sed, SQL, MongoDB.
SCIENTIFIC APPLICATIONS:	SAGE, MAP-REDUCE, RapidMiner, Weka, Pandas, Numpy, Scipy, NLTK, MATLAB, OpenCV,
VIZUALIZATION TOOLS:	NETLOGO, GraphViz, Gephi, Tableau-8.0, CytoScape, Pajek
WEB:	D3JS, C3.JS, DIMPLE.JS, XML, FLASK, CSS, HTML

Relevant Research Work : Machine learning, mathematical modelling

- **Relevant research work:** Applied information theory and unsupervised learning algorithms to find Fractal dimension of Genome/Proteome. Infectious disease modelling, modelling and simulation of evolution (using MATLAB, CytoScape, NETLOGO)

- **Web data scraping from KEGG:** Data was retrieved from KEGG in xml format, cleansed and presented in matrix format for each of the files.

- **Social network analysis:** Twitter sentiment analysis, Gephi for graph and connection visualization and segmentation. Behavior and trend analysis. Client: Memphis tech

- **Structure & Dynamics of Social Networks: Effects of perturbations on a network**

- **Language model using maximum entropy approach.**

- **PlagDetect :** Web based plagiarism detection application.

- **Network analysis for identification of hubs for knockout studies.**

- **Stochastic model for Bird flock simulation, population dynamics. Time Series Analysis.**

- **Application of NCD as universal clustering technique:** Using Information Theoretic approaches.

- **Analysis of the complete metabolome using NCD (With Dr. Tiratha Raj Singh, JUIT)**
  Graph representation of all the metabolic networks in order to detect key insights and patterns.

- **Mathematical modelling of pathways for increased production of Co-Q10 (Ubiquinone),** At Piramal Life Sciences **-- Selected for presentation at IIT, Delhi, 2nd Asian Congress on Biotechnology (ACB2013)**
  - Manuscript under preparation

- **Correlation Dimension of protein space using NCD (Normalized Compression Distance).**
  - Manuscript under preparation

Highlights of Qualifications

- Batch topper throughout my M.Tech. with a C.G.P.A. of 8.9.
- Recognized for excellent analytical skills and complex data analysis.
- Developed program for predicting solvent exposed regions in Protein-Ligand complexes
- Porting genome, proteome as well as metabolome data to the MapReduce/Hadoop framework.

Publications

- **Prediction of ligand binding sites in RNA binding protein pockets using Support Vector Machines. – Journal of Biomedical Sciences , doi : 10.3823/1020**

We presented here a SVM based approach for successful prediction of RNA-binding protein pockets. The method employs two datasets: the protein sequences of the RNA binding & non_RNA binding protein pockets. Analysis was applied on 3 different featured datasets viz FPocket, Zernike & shell features. The results suggest that the top 10 features of shell play a pivotal role in the classification & prediction of ligand binding sites in RNA binding proteins & an accuracy of 89.3% was achieved when evaluated.