

Stuti Agrawal

A dedicated professional who continuously learns and strives to find insightful results using data driven approaches.

📍 555 W. Madison St. Apt-2510
Chicago, IL 60661
☎ +1 (412) 692-0529
✉ stuti.agrawal@gmail.com
in <https://www.linkedin.com/in/agrawalstuti>

WORK EXPERIENCE

FEB 2014 – PRESENT

Center for Data Intensive Science
University of Chicago, Chicago - Illinois
Research Specialist

• Design & Development:

- Single-handedly developed computationally scalable data processing pipelines for harmonization, quantification and quality control of large scale next generation sequencing data for the **Genomic Data Commons** project (> 4 Petabyte).
- This project was released in June, 2016 by Vice President Joe Biden – an initiative of the cancer moonshot program of the United States Government.
- Engineered and productionized the ETL workflows for the above pipelines using Simple Linux Utility Resource Manager (SLURM), a highly scalable workload manager.

• Application & Analysis:

- Proactively researched, computed and validated pathological misdiagnosis on >1,000 lung cancer patients using statistical and machine learning models (**PCA, K-means, SVM**) on the harmonized data.
- This analysis formed the center's key highlight to demonstrate the applications of big data in the field of precision medicine.

• Data Visualization:

- Rendered intuitive visualizations of the results from these models in an easy to navigate web-application framework (RShiny) for facilitating research projects.

• Other Accomplishments:

- Adaptively evolved with new technology such as Docker containers, Common Workflow Language (CWL), AWS and R and applied them to match the project requirements.
- Actively contributed to different partner projects (International Neuroblastoma Risk Group) in addition to my primary responsibilities.

JUN 2013 – AUG 2013

DuPont Industrial Biosciences
Palo Alto - California
Intern

- Provided predictive solutions to Dupont's protein engineering department to increase the efficacy of their detergents by analyzing their proprietary datasets using **information theory**.

RELEVANT PROJECTS

JAN 2013 – MAY 2013

Carnegie Mellon University
Pittsburgh - Pennsylvania

Classification of Time Series Data

- Implemented two **non-linear edit distance kernels** for a SVM based classification of MS patients using dynamic programming on time series data.
- Improved the accuracy by upto 10% as compared to previous algorithms.

AUG 2011 – MAY 2012

Jaypee Institute of Information Technology
Noida, Uttar Pradesh - India

Undergraduate Thesis

- Architected a semi-automated text mining system to mine >90,000 abstracts and 1,000 full length articles and identify molecular interactions in obesity to aid targeted treatment techniques.
- **Jagannadham J, Jaiswal HK, Agrawal S, Rawal K (2016) Comprehensive Map of Molecules Implicated in Obesity. PLoS ONE 11(2): e0146759.**

EDUCATION

AUG 2012 – DEC 2013

Carnegie Mellon University, Pittsburgh, PA
MASTER OF SCIENCE –
COMPUTATIONAL BIOLOGY

AUG 2008 – MAY 2012

Jaypee Institute of Technology, Noida, India
BACHELOR OF TECHNOLOGY
– BIOTECHNOLOGY

COURSEWORK

10-601	Machine Learning
36-225	Probability and Statistics
03-712	Computational Methods for Modeling and Simulation
02-714	String Algorithms
15-122	Imperative Computation
02-710	Computational Genomics
03-711	Computational Molecular Biology and Genomics
15-112	Fundamentals of Programming
COURSERA	Exploratory Data Analysis; Programming in R; Statistical Inference

SKILLS

PROGRAMMING LANGUAGES	Python, R, CWL
TECHNICAL SKILLS	Docker, Linux/Unix, PBS/Slurm, RShiny, AWS, Postgres
MISCELLANEOUS	Improv Comedy, Graduate - Second City, Chicago IL