

# Parag Agrawal

Email: [paraga@cmu.edu](mailto:paraga@cmu.edu), [paraga@andrew.cmu.edu](mailto:paraga@andrew.cmu.edu), [a.parag@alumni.iitg.ernet.in](mailto:a.parag@alumni.iitg.ernet.in), [parag1102@gmail.com](mailto:parag1102@gmail.com)

\*GitHub Profile: <https://github.com/parag1102>

LinkedIn Profile: [www.linkedin.com/in/paragagrawalcmu/](http://www.linkedin.com/in/paragagrawalcmu/)

## Education

<b>Carnegie Mellon University, Pittsburgh</b> M.S. (Language Technologies Institute, School of Computer Science)	2013-2015	<b>GPA - 4+/4.0</b>
<b>Indian Institute of Technology (IIT), Guwahati</b> B.Tech. (Computer Science & Engineering)	2009-2013	<b>CPI/CGPA - 9.49/10</b> <b>Department Rank- 2 /70</b> <b>Institute Rank-4 /484</b>

## Work Experience (BloomReach, ESPN, Goldman Sachs, Google, RedHat, FAU-Germany)

### Improve Search Quality and Feature Extraction

**BloomReach:** Data Scientist

May-August, 2014

- Improved **search quality** from **39.8% to 49.7% (10% jump)**.
- Reduced** human intervention **cost** (active learning) by **USD 150K per month**.
- Used Language Models (HMM & Decision Trees) for feature extraction and Ordinal Regression (Pranking) for ranking.

### Sports Data Mining: Learning from Sports Text Commentary

**ESPN:** Insights Consultant, Machine Learning & Algorithms

May, 2013-Present

- Applied Correspondence Analysis on the sports commentary to capture the patterns seen specifically in cricket, soccer & basketball.
- Pointed out the vulnerabilities in a player's/team's strategy and thus can be used to improve the player's/team's performance.

(PAPER in ACM SIGKDD-2013 conference)(Winner of Innovative Project Award, INAE-2013, Top 5 Projects all over India)

### Pattern Recognition for Identifying Email Phishing and Insider Trading Practices

**Goldman Sachs Strategies:** Quantitative Analyst, Strategies and Trading

Jan-July, 2013

- Used clustering and collaborative filtering for identifying patterns.
- Used Decision trees based classification.

### Logging framework for Android and Google AppEngine for sending Logs to Google BigQuery\*

**Google:** Distributed Systems, Google Chrome Team

May-July, 2012

- Developed a distributed custom-logging handler to forward logs to BigQuery.
- Created AppEngine and Android client library in Python as well as Java for the same.
- Optimized the billing cost using Local Caching of logs to local disk and Google DataStore.

### Private Cloud for a corporate organization

**RedHat (Linux):** Team of 5 persons (4 Redhat Employees and myself)

May-July, 2011

It included the following technologies:

- LDAP
- Virtualization and Live Migration (using NFS)
- Clustering (using GFS)

### Parallel Multigrid Methods

December, 2011

**University of Erlangen-Nuremberg, Germany:** Mentored by Prof. Ulrich Rüde,

Delivered a 2<sup>1</sup>/<sub>2</sub> hr lecture at the Indo-German Winter Academy held in New Delhi, explaining Parallel Multigrid Methods to the faculty and students of University of Erlangen-Nuremberg, Germany and all the IITs.

Link to the presentation: <http://www.leb.eei.uni-erlangen.de/winterakademie/2011/report/content/course02/0220.htm>

## Teaching Experience (Lectures)

(Machine Learning, CMU)

<b>Big Data Analytics</b>	11-676 (Graduate)	Spring Semesters, Carnegie Mellon University
<b>Competitive Engineering - IBM Watson</b>	11-792 (Graduate)	Spring Semesters, Carnegie Mellon University
<b>Big Data Systems</b>	11-675 (Graduate)	Fall Semesters, Carnegie Mellon University
<b>Design and Engg. of Intelligent Info. System</b>	11-791 (Graduate)	Fall Semesters, Carnegie Mellon University

## Major Courses Taken

(CMU & IIT)

- |                                 |  |                               |
|---------------------------------|--|-------------------------------|
| ▪ Information Retrieval         | ▪ Language Modeling and Statistics       | ▪ Machine Learning Advanced   |
| ▪ Machine Learning and Robotics | ▪ Machine Learning for Signal Processing | ▪ Natural Language Processing |

## Key Academic Projects

(Machine Learning)

### Face recognition using PCA and LDA\*

July-Nov 2012

**Project Guide:** Dr. Sanasam Ranbit Singh, Machine Learning Department, IIT-Guwahati

Eigen face detection for the ORL database of faces, using PCA and then subsequently followed by multi-class LDA.

## Skills

<b>Languages</b>	:	C/C++ (proficient), Python, Java, Haskell, Prolog, Objective-C, XML
<b>Other Technologies</b>	:	Hadoop, Elastic MapReduce, Cascading, Mahout, Matlab, R, UIMA, Lucene, Weka