

## Interests

Algorithms, Machine Learning, Statistical Analysis, Distributed Systems.

## Education

2005-2007	M.Tech. (CS/ IT), <a href="#">IIT Bombay</a>	CPI: 9.6/10.0
2001-2005	B. Tech. (ICT), <a href="#">DAIICT</a>	CPI: 3.66/4.0
2000-2001	Class XII, CBSE	92.0% (PCM: 98.0%, Maths: 99%)

## Work Experience/ Projects

**Komli Media Inc.**<sup>1</sup>, India: Software Engineer Analytics (Dec 2008 – present)

*Machine Learning, Statistics, Map Reduce, Databases, Scalability.  
Java, Distributed Systems*

- [ViziSense.com](#): It publishes real-time online media usage and audience level statistics. It uses a combination of sophisticated statistical techniques and best practices to provide the accurate web audience measurement. I am responsible for the statistical models used to learn these numbers from online behavior of a panel of around 50,000 internet users.
- **Audience Targeting Platform**: This project involves serving ads to online users based on their demographics and interests. We predict the demographics based on the users browsing behavior. We have made various predictive models using techniques like Regression, SVMs, Clustering, etc. We have also designed a distribute architecture for scalability.
- **Ad View**: This product is designed for giving a snapshot of online advertisement industry. We report statistical data of various advertisers, publishers and ad networks. It will involve analysis of reach and frequency of online ad campaigns. I am responsible for the model as well as its design and development.
- **Keystroke based Biometric System**: In this project I am developing a biometric system for indentifying users based on their keyboard logs. This system will be used to uniquely identify number of users using a given computer. Various machine learning techniques like clustering, principal component analysis, feature selection were implemented.

**Yahoo! Research Labs, India**: Research Engineer (July 2007 – Dec 2008)

*Machine Learning, Recommendation Engines, Map Reduce, Hadoop, PIG!  
Java, C++, Perl, Shell, Distributed Systems.*

- **Recommendation Engines**: Worked on a collaborative filtering based recommendation engine for an advertisement platform. It recommends relevant ads for a given web page, using a stochastic model. I have single handedly designed and implemented a scalable algorithm using map-reduce/ distributed paradigm. It can efficiently handle repository of over million ad and web pages. Similar model is also extended for other Yahoo! properties like Shopping, Travel etc where we recommend different items based on user characteristics and their previous purchase record. <http://www.stanford.edu/group/mmds/slides2008/agarwal-d.pdf>
- **Opinion Reconciler**: Designed an opinion reconciler, which makes consensus on the information extracted from a common data source by multiple extraction models. A probabilistic model was designed for this based on link analysis algorithms like *pagerank* and *HITS*.
- **Linear Algebra Library**: I worked on building scalable machine learning libraries, based on distributed systems. Currently my focus is on developing a matrix algebra library based on map reduce paradigm. The philosophy behind the project is to use MapReduce and Hadoop as frameworks for distributed computing on multi-core machines, and make the computations for large datasets more than twice as fast.  
*Reference: [Dr. Deepak Agarwal](#), [Dr. SrujanaMerugu](#)*

<sup>1</sup> Komli Media was chosen among Global 50 top startups by Business Week/ YouNoodle in June 2009.

- **Emerging Technologies Group, Reliance India:** Intern (Jan-July'05)  
*Probability, Bayes Models, HMM, Perl*
  - **Machine Understanding of Indian Spoken Languages:** In this project, we designed algorithms for modeling language recognizers, based on Bayesian analysis of phoneme sequences. An attempt was also made to design an algorithm for extracting phoneme sequences in form of more abstract classes derived by statistical tools like HMM.  
*Reference: [Dr. JitendraAjmera](#), Dr. SandeepSibal*
- **National Security Depositories Ltd (NSDL), India:** Consultant(July'06 – Feb'07)  
*Network Security, VLANs, Firewall, NAT, Security Tools*
  - **Security Analysis of Very Large Networks:** In this project, we did security assessment of a large network infrastructure of National Security Depositories Ltd (NSDL). It also involved a survey of various network security tools like Nessus, RetinaScan, etc. NSDL provides online interface for tax filling, which is expected to be visited by millions of users every year.  
*Reference: [Prof. Om Damani](#), Prof. Sridhar Iyer*

## M.Tech. Thesis

### Scalable infrastructure for index based information extraction over large document collections:

*Information Extraction, Search Engine and Indexes, CRF, Data Structures.*

In this project, we designed efficient algorithms for extracting structured entities using inverted indexes over large document collections. Extraction model was based on tools like CRF. We ultimately aim to come up with index structures, which search engines can use to provide specialized search option.

*Guide/ Reference: [Prof. Sunita Sarwagi](#), [Prof. Soumen Chakrabarti](#)*

## Invited Talks

- On **Data Mining and Mining support in SQL server and Oracle**, in CEP Advance DBMS course for National Informatics Corporation (NIC), 2006
- On **Pig!and implementing large-scale ML algorithms**, Yahoo! Research, 2006

## Other Projects

- **CRF based Acronym Finder for web documents**(HyperText Retrieval and Mining): In this project we designed a CRF based tool for extracting acronyms from web documents. Aim was to explore CRF as an extraction tool.*(Java, CRF)*
- **Progress Estimator for SQL queries in POSTGRES:** Aim of this project was to make a query progress estimator using initial estimates given by the optimizer and then fine tuning it along with the query execution, based on various runtime statistics and heuristics.*(Language/Tools: C, DDD)*
- **ArchEd : A Software Architecture Based Design and Evaluation Tool Suite :** In this project we developed an eclipse plug-in to provide visual editing capabilities to monitor software life cycle based on Software Architecture concepts. During the course of this project, we also explored Eclipse Modeling Framework (EMF) and Graphical Editing Framework (GEF).*(Client: Software Engineering Institute, CMU, PA) (Java/eclipse)*

## Programming Expertise and Other Qualifications

Operating Systems	UNIX, Linux, Windows
Programming Languages	Java, C++, C, PHP,JSP,Perl, Shell scripting
Databases	Oracle, My SQL, POSTGRES
Open Source Tools modified	WEKA(JAVA), POSTGRES(C), MG4J(JAVA), Eclipse Plugins
Tools/ Editors	Matlab, Eclipse, Latex, Weka, JavaCC, CVS/SVN, Ant, JUnit
Other Qualifications	Cisco Certified Network Administrator (CCNA 2, Score 950/1000)

## Major Academic Honors

- Ranked **1st** in a class of 55 students of IT Dept (**Sem 2, SPI 10.00**), IIT Bombay
- Ranked 5th in the institute in a batch of 240 students, DA-IICT Gandhinagar

## Extra Curricular Activities

- Member of Campus Committee, Yahoo! 2008. My team was responsible for recruitment of around hundred students from colleges across India.
- Department Representative in Placement Committee, IIT Bombay 2006. My team was responsible for placement of more than 1000 students from IIT Bombay.
- First prize in National Level Network Hacking competition at Synapse 04, DA-IICT
- Inter College Dance Champion, at Syanpse-04, DA-IICT.
- Member of Cricket, Basketball, Football, and Badminton teams in college.