

**INTERESTS:** Statistical Machine learning, Data-representation, NLP, Cognition, Empirical inference, Data-mining applications.

**EDUCATION:** BTech-MTech Dual degree  
Department of Computer Science and Engineering  
Indian Institute of Technology Kanpur  
Graduate CPI: 9.6/10, UG CPI: 8.8/10  
GRE Verbal: 800 Quant: 800

**AWARDS**

- Thrice awarded 'Academic Excellence award' by the institute for distinctive academic performance (2005-06, 2007-08 and 2008-09).
- Awarded National Talent Search (NTS) Scholarship by National Council of Educational Research and Training (2002).
- Highest Marks in India in Maths and Science in the International Assessment for Schools by University of New South Wales, Australia (2 gold medals) (2001).
- All India Rank 11 in the Second National Science Olympiad, 2001.
- Twice awarded medals for proficiency in Maths in National Talent Search Competitions.
- Awarded scholarship by Council of Science and Technology, UP for highest marks in Science in school leaving exam.

**EMPLOYMENT** **Quantitative Analyst, Tower Research Capital LLC**  
(July 2010 onwards)  
— Designing and developing algorithms and programs for algorithmic high frequency trading of diverse assets at major financial exchanges.  
— Role includes formulating strategies, and developing statistical models for predicting short duration market movements.  
— Analyzing market data to identify dynamic trends .  
— Independently designing and creating system for dynamic online statistical models, instead of static ones.

**RELEVANT PROJECTS**

**Master's Thesis:** Statistical models of Language Evolution in Multiple agent communities  
— Analysis of social interaction and language learning models, for different community graphs, extending a well known syntactic model of language.  
— Analyzed language mingling and divergence in communities by stochastically coupling agent based approaches with population dynamics.  
— Identified the role of social topologies in shaping the course of linguistic evolution and language divergence.

Guide: *Dr. Harish Karnick*, IIT Kanpur

**Predicting click-rates of online Advertisements**

— Industry project by Komli to identify attributes and patterns which can lead to higher click rates for advertisement on websites.  
— Use of local clustering for rare class analysis, and SVMs for classification.  
— Adjudged as **best project** among more than 20 proposed solutions.

Guide: *Dr. Harish Karnick*, IIT Kanpur

**Biocomputing:** Using Gene Ontology contexts for Microarray predictions

— Concise representation of microarray data in terms of biological concepts, using a knowledge infusion from the GO database, and extending the statistical approach.  
— Prognostic predictions corroborate literature, and identify precise biochemical pointers.

— Poster presented at Asia Pacific Bioinformatics Conference, 2010.  
Guide: *Dr. Arnab Bhattacharya*, IIT Kanpur

### **Autoguider for Astrophotography**

*Internship at Max Planck Institute Tubingen. May-July, 2009*

— Designed a two variable Q-learning type formulation for autoguiding a Mach GTO German Equatorial Mount for star-tracking and astrophotography.

— Formulated problem as a continuous MDP with agent ‘actions’ corresponding to motor movement durations, and fine tuned simultaneous learners for Right Ascension and Declination axes.

— Indoor simulations using convolutions on laser point source for robust star detection.

Guides: *Dr. Bernhard Scholkopf, Dr. Jan Peters*

### **Data Mining:** A Similarity search in Monophonic melodies

— Sequential use of global feature based clustering and matching transportation distances to return  $k$  most similar melodies to sample query from an unlabeled music database.

— Extended to inputs from a Midi keyboard; encouraging results on two standard datasets with amateur players.

— Poster presented at IBM Industry-Academia workshop at New Delhi.

Guide: *Dr. Arnab Bhattacharya*, IIT Kanpur

### **Activity Prediction with Physiological Time series data**

— Comparative analysis of classification algorithms on the Physiological Modeling Dataset, with emphasis on the performance of Support Vector Machines.

— A semi-supervised voting-based approach beat the benchmark of this task in the ICML-2004 Physiological Data Modeling Contest.

Guide: *Dr. Harish Karnick*, IIT Kanpur

### **NLP:** Disambiguating Hindi Named Entities

— Analyzed the role of associated Parsargs, Verb Classes and multiple sources of evidence. Identified role of parsargs and POS tags as critical in Hindi NER.

— Ensemble learning approach, using Cost sensitive learning, achieved classification accuracy of 77%, and best F-measure of 0.56, improving both accuracy and recall.

— Hybrid **system integrated** for NER at Language Lab, IIT Kanpur.

Guide: *Dr. RMK Sinha*, IIT Kanpur

### **UCSD Data Mining Contest:** Credit Card Fraud Detection

— **Best Indian team** (UG and PG) at UCSD 2009 Data Mining Contest.

— Detected 85.7% of fraudulent transactions (lift of 4.286 at 20% ) using J48 Decision trees and extensive feature preprocessing.

## **TALKS AND PRESENTATIONS:**

- Departmental seminar on ‘graph clustering’ in SIGML group, IIT Kanpur. Sep 25, 2009
- Invited talk and practical workshop on ‘Introduction to Machine Learning’, Geetanjali Institute of Technology and Science, Udaipur, in 2010
- ‘A GO based representation for prognosis and inference from microarray data’, poster presentation at Asia Pacific Bioinformatics Conference (APBC), 2010

## **RELEVANT COURSES:**

**Graduate:** Machine Learning and Knowledge Discovery, Advanced Graph Algorithms, Statistical and Algorithmic methods in Bioinformatics, Special Topics in Computer Science (independent study), Indexing and Search Techniques, Advanced NLP, Parallel Algorithms.

**Undergraduate:** Linear Algebra, Design and analysis of algorithms, Data structures and algorithms, Theory of Computation, Special Topics in Computer Science (independent study), Databases, Computer Networks, Compiler Design, Operating Systems, Computer Organization, Software Engineering, Principles of Programming Languages, Real and complex analysis, Microeconomics, Differential equations.

**TECHNICAL  
SKILLS:**

Languages : C, C++, Matlab, Perl, Shell, Java  
Software : Eclipse, WEKA, RapidMiner  
Web Technologies : PHP, HTML, Javascript  
Platforms : Linux, Windows  
Other : MySQL, Make, Awk, Lex, Tex, Yacc and Shell Programming.

**OTHER:**

- Head coordinator, Summer school for Computer Literacy for rural children, as member of 'Shiksha Sopan'(regional NGO).Initiated and synchronized practical computer training through 8 secondary tutors, and extended course in villages of Barasirohi and Nankari.
- Senior Link Student and Student Guide, Counseling Service IIT Kanpur: assisted all 5 allotted counselees (highest in year among all link students) to recover from academic warning, and responsible to provide emotional and academic support to eight students throughout campus stay.
- Selected to attend 'Microsoft Winter School on Machine Learning and Computer Vision' at IISc Bangalore (Jan 2010).
- Teaching Assistant for 'Data Structures and Algorithms' for Spring Semester, 2009; and for 'Machine Learning and Knowledge discovery' (Fall semester, 2010).