

Research Interests:

Machine Learning, Natural Language Processing, Information Retrieval, Speech Recognition

Education:

Master of Science, Computer Science and Engineering (2013-2016) IIIT Hyderabad Advisor: Dr. Vasudeva Varma, Professor and Dean Research, IIIT Hyderabad	8.64/10
B.Tech in Computer Science and Engineering (2009-2013) Heritage Institute of Technology Kolkata, West Bengal University of Technology	8.81/10
Higher Secondary Examination (2007-2009) Nava Nalanda, West Bengal Council of Higher Secondary Education	86.4 %
Secondary Examination (2006-2007) Gokhale Memorial Girls' School, West Bengal Board of Secondary Education	92.6 %

Experience:

Apr'17-Present	Data Scientist at Bing, Microsoft AI+R. Working on Automatic Speech Recognition for Indic Languages mainly targeting Cortana, Microsoft Cognitive Services and other client projects. Working on acoustic modeling - improving performance for Indian English and Hindi, bootstrapping and shipping models for other regional languages as a first.
Aug'16-Mar'17	Software Development Engineer at Bing, Microsoft India RnD Pvt. Ltd. Worked on improving contextual relevance and ranking for Bing search.
Mar'16 - Jul'16	Research Intern at Text and Graph Analytics Group, Xerox Research Centre India.
Jan'14 - Jul'16	Research Assistant at Search and Information Extraction Lab (SIEL), IIIT Hyderabad.
Apr'15 & Jul'14	Data Science Intern at Flipkart
Jan'15 - May'15	Project Mentor for Information Retrieval assisting Dr. Vasudeva Varma.
Aug'14 - Dec'14	Teaching Assistant for Web Mining assisting Dr. Manish Gupta.
Jun'12 - Jul'12	Summer Intern at Center for Data Engineering, IIIT Hyderabad guided by Dr. Kamal Karlapalem
Jul'11 - Jul'11	Summer Intern at Department of CSE, IIT, Kharagpur guided by Dr. Rajib Mall.

Publications:

- **TweetGrep: Weakly Supervised Joint Retrieval and Sentiment Analysis of Topical Tweets**, ICWSM 2016, Cologne, Germany
Satarupa Guha, Tanmoy Chakraborty, Samik Datta, Mohit Kumar, Vasudeva Varma
- **SIEL: Aspect Based Sentiment Analysis in Reviews**, SemEval 2015, Denver, Colorado
Satarupa Guha, Aditya Joshi, Vasudeva Varma
- **Sentibase: Sentiment Analysis in Twitter on a Budget**, SemEval 2015, Denver, Colorado
Satarupa Guha, Aditya Joshi, Vasudeva Varma
- **Hardware Acceleration using Mitrion-C**, Inter-Engineering College Academic Meet 2012, Kolkata, India
Satarupa Guha, J. M. Manasa

Patents:

- Performing Meaningful Semantic Analyses Of User-generated Textual And Voice Content
Application no. 20160645US01, done as a part of internship at Xerox Research

Skills:

Programming Languages	Python, C#, Java, Matlab, C
Web Languages	HTML, CSS, Javascript
Libraries and Tools	Numpy, Scipy, ScikitLearn, NetworkX, Vowpal Wabbit, LibSVM, CNTK, Keras Stanford coreNLP, Lucene, Elastic Search, Mallet, MySQL, Bash, Git, L ^A T _E X

Relevant Course Work:

Graduate Level Courses

- Machine Learning • Natural Language Applications • Information Retrieval
- Statistical Methods in AI • Web Mining • Data Mining and Warehousing

Undergraduate Level Courses

- Data Structures and Algorithms • Artificial Intelligence • Image Processing
- Operating Systems • Database Systems • Networking

Databases by Jennifer Widom from **Coursera** (Certificate of completion)

Awards and Achievements:

- **AIR 780** out of 224160 in the paper Computer Science and Information Technology in GATE 2013 (top 0.25 %).
- **Reviewer** for Natural Language Conference ICON 2015,2017.

Select Research Projects:

Learning spatio-temporal and semantic patterns from social media (*Mar 2016 - Jul 2016*)

Working with Manjira Sinha, Preethy Varma and Tridib Mukherjee at XRCI, in order to identify, track and analyze location-specific governance issues over time, so that public grievances can be addressed as efficiently as possible.

Weakly Supervised Joint Retrieval of Topical Tweets and Sentiment Analysis from Twitter (*Jan 2015 - Jan 2016*)

Worked with Samik Datta, Dr. Tanmoy Chakraborty, Dr. Mohit Kumar and Dr. Vasudeva Varma on a weakly supervised approach for joint retrieval of topical tweets as well as the associated sentiment for the topic. The novelty of the approach is two-fold: we augment the retrieval of topical tweets using prior information about sentiment around the topic as well as improve the sentiment analysis by utilising the additional context of the topic. Alleviating the need for tedious manual annotation, this approach uses supervision in the form of a few keywords related to the topic and expected label proportions, following the semi-supervised Expectation Regularization paradigm. (*Published at ICWSM 2016.*)

A Twitter Conversation Assistant (*Apr 2015 - Jan 2016*)

Worked with Dr. Manish Gupta and Dr. Vasudeva Varma on building a Twitter Conversation Assistant - a system that identifies the tweet most likely to get a reply in a conversation and suggests a reply to the tweet based on the context of the conversation and the content of the tweet. The work involved building regression models using weak labels for millions of tweets.

Aspect Based Sentiment Analysis in Reviews (*Oct 2014 - Jan 2015*)

Worked with Dr. Vasudeva Varma to identify the aspects of entities in a document and the sentiment expressed for each aspect of those entities. (*Published at SemEval-2015.*)

Sentiment Analysis in Twitter (*Dec 2014 - Jan 2015*)

Worked with Dr. Vasudeva Varma on building state of the art system for analyzing sentiment in tweets. The work mainly focussed on feature engineering and supervised learning. (*Published at SemEval-2015.*)

Detecting Influential Nodes for Reachability in Social Networks (*Jan 2013 - Jan 2014*)

Worked with Prof. Partha Basu Chowdhuri, on a problem which can be generalized as k-hop dominating set problem, where a maximum of k hops will be allowed to spread the information among all the nodes of the graph. (*under review at the Information Processing and Management Journal.*)

Determining Maximal Common Subgraphs in Geometric Graphs (*June 2012 - July 2012*)

Worked with Dr. Kamal Karlapalem, to find similarities in the underlying graph representations of a set of geographic locations. It is essentially an application of graph isomorphism problem in order to find similar road structures in different cities from OpenStreetMaps.

Extra Curricular Activities:

- Volunteer at Asha Kiran, school for under-privileged children at IIIT.
- Basic training in Indian classical music.
- Basic training in swimming.