Mansi **Gupta**

Applied Research Engineer · LinkedIn

(+91) 8971847474 | mgupta1410@gmail.com | www.mansigupta.in | mgupta1410 | mansiguptain

Research Objective

I am interested in the areas of **Machine Learning** and **Natural Language Processing** with the focus on addressing **global social problems**. My work has involved behavioral analysis and influence propagation on social networks, recommendation systems and information retrieval on large corpora. I often wonder about fairness and interpretability of ML algorithms.

Education

Birla Institute of Technology and Science (BITS) Pilani, Pilani Campus *Aug 2011 - Jul 2015*

Bachelor of Engineering (Hons.), Computer Science (GPA 8.47 / 10) *Pilani, India*

- Undergraduate thesis on 'Near Duplicate Detection in articles' at *LinkedIn*, India

Experience

Research

Applied Research Engineer *Jul 2016 - Present*

Spam and Relevance, LinkedIn (Advisor: **Dr. Alpan Raval**, **Dr. Anirban Dasgupta**) *Bangalore, India*

- Predicting tendency of LinkedIn members to produce and spread unprofessional content (**Spam Reputation**) on the network
  - Modeled the prior scores using gradient boosted trees with supervised scores computed using labels on content shared by users
  - Analyzed community structure with respect to spread of low quality content on the graph with ∼**33 million** nodes and ∼**4 billion** edges
  - In process of developing a semi-supervised parametric **label propagation** algorithm to account for missing data and propagate influence
- Experimenting with different autoregressive topic modeling techniques to detect **topic change points** in transcripts of Lynda videos

Research Intern *Jan 2015 - Jun 2015*

Search, LinkedIn (Advisor: **Dr. Chandramouli M**) *Bangalore, India*

- Implemented **Near Duplicate Detection** in LinkedIn articles to weed out plagiarized content with **precision of 97%** and **recall of 74%**
  - Researched and compared various techniques for near duplicate detection on large corpus having ∼**11 million** documents
  - Implemented **SpotSigs** algorithm to convert document to sets, **Minhashing** to generate document signature
  - Reduced time complexity of comparing pair of documents from $O(n^2)$ to $O(k * n)$ by using **Locality Sensitive Hashing** to cluster likely similar documents into k clusters

Independent Research

- Worked on the problem of network based personalization for interest prediction in large social networks **Link**
  - Proposed **Network-aware Determinantal Point Processes** for selecting representative yet diverse nodes from a user's network
  - Modified Nyström's approximation to reduce dimensionality by adaptively selecting reputed users as landmarks instead of fixed landmarks
  - Achieved a gain of **27% in precision@15** and **10.3% in recall** over the baseline of using the entire network

Freelance Researcher and Content Writer *Oct 2016 - Present*

SocialCops *Delhi, India*

- Performed structural and textual analysis on speeches of the Indian P.M. Mr. Modi to uncover issues addressed over time **Link**

Industry

Software Engineer *Jul 2015 - Jun 2016*

Search, LinkedIn *Bangalore, India*

- Developed features and enhanced relevance for search on different verticals like articles, universities and companies
- Worked on **entire lifecycle of a search query** - detecting query intent, entity resolution, indexing, retrieval and personalized ranking
- Built search for **LinkedIn Learning** including relevance for autosuggest and search page
- Introduced **'Statistically Important Phrases'** in LinkedIn Learning search which led to the increase of **7%** in Click Through Rates **(CTRs)**
- Formulated query rewriting modules to handle complex queries on **Codesearch**, an internal search tool on LinkedIn's codebase
- Troubleshooted production issues like load distribution and handling live updates while each node is serving ∼**600 queries/sec** live

Network Engineering Intern *May 2014 - Jul 2014*

Infrastructure and Operations, LinkedIn *Bangalore, India*

- Automated the process of monitoring and processing alerts due to issues in physical links in LinkedIn's backbone network
- Led to reduction in time spent by Network Engineers in manual monitoring of alerts by ∼**75%**

Data Science Intern *Jun 2015 - Jul 2015*

Access Healthcare International *Delhi, India*

- Computed correlations in factors affecting **Maternal Mortality** in the state of Madhya Pradesh to develop **Theory of Change** flow **Link**

Publications

[1] Rishabh Mehrotra, Mansi Gupta (2016). Network-DPPs: Exploiting User's Network for Interest Prediction. Manuscript in preparation

Academic Projects

Broad Topic Intention

Study Oriented Project

- Studied various statistical and semantic language models for retrieving documents given queries with broad intention
- Implemented **conceptual query model** and traditional term-based model on TREC Genome collection
- The conceptual model outperformed by increase of **6.4% in MAP** (Mean Average Precision) and **4.2% in Precision@10**

Topic-wise Influence Mining and Reach Estimation on a subgraph of Twitter

Link

- Determined the influence of users with respect to specific topics (like Politics, Sports, etc) on Twitter
- Classified tweets using **multinomial Naive Bayes** and measured influence of a users using **Information Cascade Model**

Online Clustering of Parallel Data Streams

Link

- Clustered transient sequence of time-stamped values on the basis of their evolution over time, using a **parallel version of K-Means** algorithm
- Improved time complexity from $O(n^2)$ **to** $O(n \log n)$ by using **Haar 1D Wavelet transforms** as opposed to Discrete Fourier Transform

Parallel implementation of sequence alignment of nucleotides

- Developed parallel version of **Smith-Waterman-Algorithm** for sequence alignment in CUDA, achieved speed up of **5-6 times** on 48-core GPU

Parallel Implementation of Page Rank algorithm

Link

- Implemented Page Rank algorithm on a cluster of machines. Compared Shared Memory Model, Message Passing Interface and their hybrid

Class Performance Evaluation Tool

- Predicted the probability of getting a question correct in GMAT or SAT given its categories, achieved **73%** accuracy using **decision trees**

Designing Efficient Datacenter topology

Study Oriented Project

- Simulated a **degree-bounded random graph topology** which outperformed traditional fat-tree topology by supporting **27% more servers**

Teaching and talks

Co-founder, GetPlaced()

- Organized 7-days long workshops in three campuses of BITS Pilani to help students prepare for recruitments in Computer Science jobs
- Delivered lectures covering advanced data structure and algorithms. Set up and conducted mock interviews and programming tests

Presentations

- Invited at **IIT Gandhinagar** to address an audience of researchers, professors and students on the topic **'Deep dive into Search Systems'**
- Invited by a company, WeBind, to address students of five engineering colleges in a **webinar** on the topic **'Basics of Search Systems'**
- Addressed Software Engineers and researchers on internal workings of **Galene, LinkedIn's search infrastructure** at LinkedIn
- Presented the seminar on **'Near Duplicate Detection'** to Software Engineers at LinkedIn
- Presented and discussed various papers at the **Machine Learning Reading Group** at LinkedIn
- Delivered the **farewell speech** on the behalf of 2015 batch of BITS Pilani at the Director's tea party

Extra curricular Activities

Volunteer, Nirmaan Organization

Aug 2011 - Dec 2013

- Worked as a volunteer in a Self Help Group of women of disadvantageous community to help them gain financial independence

Coordinator, Alumni Research Talks

2013 - Present

- Organizing a research symposium from four years now, where BITS Pilani alumni and faculty speak on research topics in Computer Science

Problem Setter

- Set problems for **Codestorm**, an intercollegiate coding competition in BITS Pilani, with participation of 100+ teams *Jan 2014*
- Set problems for **Wintathon**, a women in tech hackathon organized by LinkedIn *Sep 2016*

Technical editor, Embryo Club, BITS Pilani

Aug 2011 - Dec 2014

- Compiled and edited articles related to latest research in Computer Science for a monthly newsletter

Academic achievements

- **Topper, BITSAT 2011** (BITS Pilani, Entrance Examination), attended by 120,000 candidates *2011*
- Secured **All India Rank 342** in India's Largest Competitive Exam, AIEEE 2011, attended by more than 1 million candidates *2011*
- Stood **2nd** nationwide in ACM Lady Ada Programming Contest *2014*
- Stood **28th out of 1056** teams in Women's Cup organized by HackerRank *2015*