

# PRABHAT AGARWAL

Stanford, CA 94305

(650) · 521 · 7099 ◊ prabhat.agarwal@cs.stanford.edu ◊ linkedin.com/in/agarwal-prabhat

## EDUCATION

### Stanford University

*MS in Computer Science*

**Cum. GPA:** 4.1 / 4.0

**Relevant Courses:** Artificial Intelligence (CS221), Machine Learning with Graphs (CS224W), Reinforcement Learning (CS234), Natural Language Processing with Deep Learning (CS224N), Convolutional Neural Network for Visual Recognition (CS231N), Deep MultiTask and Meta Learning\* (CS330)

2019-Expected June 2021 — Stanford, CA

### IIT Kharagpur

*B. Tech (Hons.) in Computer Science & Engineering*

**Cum. GPA:** 9.93 / 10.0 (4.0 / 4.0)

**Relevant Courses:** Machine Learning, Information Retrieval, Social Computing Speech & Natural Language Processing, Deep Learning, Operating Systems, Parallel & Distributed Algorithms

2013-2017 — Kharagpur, WB, India

## PUBLICATIONS

- [1] P. Agarwal, A. Sharma, J. Grover, M. Sikka, K. Rudra, and M. Choudhury. I may talk in english but gaali toh hindi mein hi denge: A study of english-hindi code-switching and swearing pattern on social networks. In *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*, pages 554–557. IEEE, 2017. <https://doi.org/10.1109/COMSNETS.2017.7945452>.

## EXPERIENCE

### Pinterest

*Research Intern, Pinterest Labs (Trust & Safety)*

Jun 2020 - Sept 2020

California, US

- Modeled the problem of spam detection as semi-supervised learning over the interaction graph of different spam entities to capture various interactions and clusters between the spam entities like users, pins, and links.
- Designed and developed a graph-based classification model (GraphSAGE and GAT) to jointly classify entities as spam over this graph containing 40M nodes and 240M edges.
- Achieved significant incremental filtering of weekly spam impressions (> 3% of total pin impressions) over production models.

### Goldman Sachs

*Analyst, SecDb Architecture*

Jun 2017 - Aug 2019

Bengaluru, India

- Led a team of 4 to develop a system to provide a central service for real-time queries on the firm's primary trade and risk data, with horizontal scalability, high availability, and multi-region deployment.
- Worked on developing a proof of concept for a scalable, highly available distributed database to allow trade booking and real-time risk calculation for all firm businesses with sub-millisecond read-write latencies.

## KEY PROJECTS

### Predicting Safety Of Clinical Trials

*Research Assistant, Dr. Jure Leskovec (Stanford Univeristy)*

Oct 2019 - Present

- Built a first of its kind knowledge graph (KG) of 120k clinical trials by extracting structured info from ClinicalTrials.gov and combining existing biological knowledge bases like drug-protein network, PPI, UMLS, DrugBank, etc.
- The knowledge graph contains 200k nodes, 4M edges, and over 30 relations (e.g., trial-side-effect, drug-drug) and 10 entity types (e.g., side-effect, drug, disease, population, outcome measures).
- Working on building a graph learning model to predict the side-effects of a clinical trial given the drug, disease, and other characteristics of the tested cohort of patients.

### FactRanker: Automatic Ranking Of Check-Worthy Claims

*B.Tech Thesis, Dr. Pawan Goyal (IIT Kharagpur)*

Jul 2016 - Apr 2017

- Curated a dataset of political claims annotated using all major fact-checking media outlets using a semi-automated process.
- Designed a system FactRanker to rank claims by their check-worthiness using an SVM classifier trained on text (e.g., POS tags, sentiment) and contextual features (e.g., topic, sentence homogeneity), improving the then state-of-the-art (ClaimBuster) by 21.7% in NDCG@100.

## TECHNICAL STRENGTHS

### Computer Languages

Java, C/C++, Python, Javascript, Bash, SparkSQL

### Frameworks

scikit-learn, Tensorflow, nltk, PyTorch, PyTorch-geometric, PySpark

### Others

gdb, Git, asio, Network and Systems programming