

RESEARCH STATEMENT

Ankur Sharma ([ankursharma-iitd.github.io](https://github.com/ankursharma-iitd))

30th November, 2020

It seems almost magical: From getting personalised recommendations on what posts to read on Facebook to promoting businesses on Instagram, social networks have completely redefined the way we interact with our society and technology today. I am Ankur Sharma, currently a research intern in the Social Media Analysis team (under [Prof. Joyojeet Pal](#)) of Technology for Emerging Markets group at **Microsoft Research (MSR)**, Bangalore. I have graduated with a Bachelor of Technology (B.Tech) and Master of Technology (M.Tech) in Computer Science from **Indian Institute of Technology (IIT) Delhi**. My primary research interests lie in understanding these complex human social networks, with focus on the use of new and traditional media technologies that affect our day to day lives. I wish to deploy my skills in data science with state-of-the-art machine learning and natural language processing techniques to curate and archive content, categorize it in suitable typologies and identify significant trends in the data using statistical modelling. I wish to pursue interdisciplinary research in computational social science so as to create a positive societal impact in the real world by helping people in tackling various humanitarian challenges. My recent research projects have been centred around the use of social media platforms in mainstream politics, specifically around political brand building in India. My long-term career plan is to be a researcher in applied machine learning – either in industry or in academia. Applied Machine Learning entails the need for cross-domain research on large real-world datasets that span several related areas in natural language, vision and media studies, amongst others.

My inclination towards the field of engineering had motivated me to work hard and build my fundamental concepts in physics and mathematics from an early age. Strong command over these concepts enabled me to secure an **All India Rank of 54** in the IIT Joint Entrance Examination (JEE) 2015 from among a million candidates. My ardor in Mathematics, coupled with my hands-on experience in coding with Java & C++ during high school, sparked an interest in pursuing my undergraduate studies in Computer Science. My drive to excel continued in IIT Delhi, wherein during my Master's, I was ranked among the top 5 students in my degree program with a high grade point average of 9.24 on a 10 point scale.

To decide if research would be the right fit for me, I pursued an internship at the **University of Illinois at Urbana-Champaign (UIUC)** in the summer of 2017. I was supervised by [Prof. Marianne Winslett](#) on designing a novel attack framework to successfully reverse engineer the blueprints of an object from the factory floor using magnetometer side-channel data. Exploiting the design shortcomings of a 3D printer, I was able to experimentally show that the placement of motors on such a device can render it insecure. Integrating the entire reconstruction pipeline was a significant challenge since classical methods of segmentation on magnetometer spectrograms failed to generalize to new settings and environments. To build more adaptable systems, I started exploring the field of Machine Learning, which became a pivotal point in my career.

Having a successful learning experience during the internship, I decided to pursue the field further. This was the time when I started working on a research project with [Prof. Aaditeshwar Seth](#) & [Prof. Parag Singla](#). The main objective of the project was to analyze past data of commodity prices in India for the last ten years and figure out the reasons for high fluctuations in prices while mapping them to various possibilities such as weather, hoarding, and strikes. We came up with multivariate time series models using SARIMAX and LSTMs to model these prices accurately. To improve it further, I devised an innovative method to leverage the time-delays between different pairs of markets to account for the possible correlations between them appropriately. Eventually, we were able to show in our results that it is possible to predict the reasons for high variance in prices. For validation, we got supporting facts from newspaper articles published during the time of the price hike. Not only did this project help me gain exposure in dealing with real-world articles' dataset, but it also taught me different aspects of research, including ideation of the problem, literature review and the perseverance to explore new paths with active collaboration. I got my first technical-writing experience in this project and we published our work¹ in the ACM SIGCAS Conference on Computing & Sustainable Societies (**ACM COMPASS 2019**).

¹*Price Forecasting & Anomaly Detection for Agricultural Commodities in India* - L. Madaan, **A. Sharma**, P. Khandelwal, S. Goel, P. Singla, A. Seth - Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS '19 [↗](#)

After gaining more coursework experience in the field of Deep Learning and Computer Vision, I went for a semester-long research visit at the **Massachusetts Institute of Technology (MIT)**, where I worked on through-wall Identity Recognition under the supervision of **Prof. Dina Katabi** at CSAIL. Reflections of RF signals from the human subjects were processed to train a pose-estimation model. This trained model generates a key-point confidence map which can be extracted to obtain a raw skeleton of the human body. Many motion-relevant GAIT features were extracted to uniquely model the identity of the individual. I used aggregation strategies on these extracted features for learning unsupervised representations robust to occlusions and environments. It was tested in through-wall identification scenarios for personalized and real-time sensing in smart homes as well. The stimulating environment in the lab and my interactions with fellow graduate students helped me realize the opportunities within the pursuit of further education. It was inspiring to see people working passionately on real-world problems and trying to move technology forward. The opportunity to lead weekly meetings was also a learning experience, enabling me to develop soft skills to articulate my points well.

Upon my return, I continued working with **Prof. Aaditeshwar Seth** for my Master's thesis to study the ideological biases prevalent in the Indian mass media on key economic and technology policies that affect our day to day lives. We built an end-to-end system that starts with a news article and parses it to obtain statements made by people in the article; on these statements, we applied a Recursive Neural Network based model to detect whether the statements express an ideological bias or not. Our results show that the Indian mass media is ideologically biased, typically covering pro-policy statements much more than anti-policy statements and favoring a technology deterministic viewpoint more than the other side of the discourse. Our method can serve as a basis to contrast social media self-expression by prominent people with how the mass media portrays them. This project introduced me to media studies and how ideological biases in a media platform become an important apparatus in shaping public opinions. I took the lead in building up the research question and in paper-writing, and got my first-author paper² accepted at the International Conference on Advances in Social Network Analysis and Mining (**ASONAM 2020**). While working for the project, I got interested in the computational social science domain, and I started exploring opportunities in this area.

During my time spent at IIT, I started attending talks from the Data Analytics and Research (DAIR) group - one of the famous SIGs managed by faculty members at IITD. During my free time, I like to work on the development of products and services that help and build the social community. As a part of the Assistech Labs at IIT Delhi, I created an affordable audio-aided tactile reader [🔗](#) for visually impaired people. I also launched a personalized rickshaw service named PickMe [🔗](#) within our campus to allow easy commute. I served as a teaching assistant for core undergraduate courses in computer science - Data Structures & Parallel Programming. I am also currently a teaching assistant in the course Data Structures & Algorithms, taking hour-long tutorials twice a week. In addition to taking tutorial labs, I also helped in previewing the assignments and assisting students with the course outside the tutorial hours. I found my teaching assistantship to be one of the most fruitful experiences, elevating my desire to pursue academia in the future.

After completing my Master's from IIT Delhi, I started working with **Prof. Joyojeet Pal** as part of the Social Media Analysis Team at Microsoft Research (MSR), India. Firstly, I conducted a mixed-methods quantitative study to highlight themes of conspiracy and astroturfing across 33K politicians, 300 media houses and 3K journalists in the death of a popular film star (*Sushant Singh Rajput*) in India, which created massive uproar on the news cycle and social media. Using data from Twitter, YouTube, and an archive of debunked misinformation stories, we examined the drivers and consequences of social media outrage in this case, and contextualized the findings in the Indian socio-political backdrop. Our early-stage findings³ received widespread coverage in the press⁴ and social media. The full paper is currently under review at CSCW 2021.

Secondly, to understand the political dialogue on Twitter in terms of the topical choices of tweets, we developed a novel training method which learns the embedding of a politician by considering one tweet at a time, instead of concatenation. We use these embeddings to show that the local geography and party affiliation play an influential role in deciding what politicians write on social media in India and the US. Our work is currently under review at ECIR 2021. Taking this idea further, I modelled the engagement (mentions/retweets) activity of politicians

² *Ideology Detection in the Indian Mass Media* - **A. Sharma**, N. Kaur, A. Sen, A. Seth - (ASONAM 2020) [🔗](#)

³ *Anatomy of a Rumour: Social media and the suicide of Sushant Singh Rajput* - S. Akbar, **A. Sharma**, H. Negi, A. Panda, J. Pal [🔗](#)

⁴ [Scroll](#), [The Wire](#), [India Today](#), [The Quint](#), [Mid-Day](#) and [NDTV](#)

on Twitter as a big social graph with weighted links. The learnt embeddings show a strong correlation between the retweeting activity and the party affiliation of the politician. Further, I am currently analyzing the language divide in India on social media along with the content-retweet embeddings to understand the reasons why some states and parties show higher internal engagement than others. I am also simultaneously working on a written piece, examining the taxonomy of Indian politics on Twitter on a state-by-state basis. For each state, the role of language, key politicians in that state, the relationship of various influencers and journalists with the state politicians, connectivity with the center and the way social media campaigns are organized for all the parties within the state is being discussed.

Working simultaneously on different projects has allowed me to build my research and people skills immensely. I got an opportunity to interact with, and brainstorm ideas with people from diverse academic backgrounds and research experiences, and also got a chance to learn efficient research practices. I was given the complete ownership of the project and total flexibility in defining the research problem the way I wanted. The internship has allowed me to improve my academic writing skills, and frequent presentations helped me formulate research problems, and delineate research results clearly and concisely. Specifically, I have improved upon my result presentation by employing the use of creative visualization tools and libraries like Pandas, Matplotlib and Plotly, and Charticulator, amongst others.

Besides research, I have also pursued software engineering internships at **Microsoft**, where my work included performance analysis of Azure storage systems (2018) and creating a contextual add-ins recommender for Outlook (2019). These internships allowed me to grow as a team player while working in a cross-team collaborative environment, and as an individual, when I learned to handle large codebases for developing applications that follow specific design considerations. After these development-oriented internships, I felt that my true allegiance lay with the opportunities to delve into the broader spectrum within academia.

The whole sequence of projects and accomplishments led me to discover my research interests. I firmly believe that a new synthesis of ideas and skills takes place when disciplines collaborate for research. My goal is to devise novel methods to extend the concepts of applied machine learning in complicated settings that arise in health care, security, internet of things, and natural sciences.