
EDUCATION

- **Indraprastha Institute of Information Technology** New Delhi, India
Bachelor of Technology, Computer Science and Engineering; CGPA: 9.24/10 Aug. 2016 – May 2020
- **Sanskriti School** New Delhi, India
High School; Percentage: 93.8% 2014 – 2016

PUBLICATIONS

- Aditya Chetan*, **Brihi Joshi***, Hridoy Sankar Dutta, Tanmoy Chakraborty. CoReRank: Ranking to Detect Users Involved in Blackmarket-based Collusive Retweeting Activities. In *The 12th ACM International Conference on Web Search and Data Mining (WSDM 2019)*. (Acceptance Rate: 16%, CORE2018 A*)
- Hridoy Sankar Dutta, Aditya Chetan*, **Brihi Joshi***, Tanmoy Chakraborty. Retweet Us, We Will Retweet You: Spotting Collusive Retweeters Involved in Blackmarket Services. In *The 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2018)*. (Acceptance Rate: 15%) arXiv:1806.08979 [cs.SI]
- Nishtha Madaan, Gautam Singh, Sameep Mehta, Aditya Chetan*, **Brihi Joshi***. Generating Clues for Gender based Occupation De-biasing in Text arXiv:1804.03839 [cs.CL]

* Equal Contribution

SKILLS

- **Relevant Courses:** Machine Learning, Natural Language Processing, Semantic Web, Linear Optimization, Linear Algebra, Data Mining, Probability and Statistics, Real Analysis, Analysis and Design of Algorithms, Discrete Mathematics, ODEs, Numerical PDEs, Digital Audio
- **Languages:** Python, Java, C, Bash, JavaScript
- **Tools & Technologies:** Git, SVN, PyTorch, Keras, MATLAB, scikit-learn, nltk, spaCy, numpy, maven, Spark, SparQL

RESEARCH EXPERIENCE

- **Laboratory for Computational Social Systems** New Delhi, India
Undergraduate Researcher Jan 2018 - Present
 - **Understanding adversarial collusive activities in OSNs:**
Advised by - Dr. Tanmoy Chakraborty
 - * Worked on detecting collusion on Twitter, specifically on detecting collusive retweeters via freemium blackmarket services
 - * Curated an open dataset of manually annotated users from various freemium services
 - * Proposed a novel set of features which beat the state-of-the-art supervised models for fake users and spam bot detection on the task of collusive user detection
 - * Extended the approach by developing an unsupervised and semi-supervised approach that takes into account - network features, behavioral features and topical features

Keywords: *Machine Learning, OSNs, Data Science*
- **IBM India Research Laboratory** New Delhi, India
Undergraduate Researcher Nov 2017 - Present
 - **Complex Answer Retrieval:**
Advised by: Dr. Sumit Bhatia
 - * Working on the problem of **Complex Answer Retrieval**
 - * Performed a thorough literature survey on the existing methods employed in question answering systems
 - * Currently exploring and implementing basic techniques in Information Retrieval to improve the performance of downstream Neural Q&A models
 - * **Keywords:** *NLP, Complex Answer Retrieval, IR, Deep Learning*

- **Persuasive Style Transfer:**

Advised by - Dr. Deepak Padmanabhan (QU, Belfast)

- * Modelled the “persuasiveness” quality in political speeches
- * Wrote a linear program to make alterations to text to reduce the “persuasiveness” metric

Keywords: *NLP, Machine Learning, Deep Learning*

- **Occupational Debiasing [code][preprint]:**

Advised by - Dr. Sameep Mehta

- * Developed a system for detecting Occupational Gender Bias in text, taking into account the influence of demography and time-frame
- * Developed and curated a dataset comprising of various occupation names and occupational evidences
- * Proposed a pipeline that detects potential gender bias in occupations
- * **Best Poster Runner’s up award at GHCI 2018**

Keywords: *NLP, Machine Learning, Flask*

- **Interdisciplinary Lab for Interactive AV Development (ILIAD)**

New Delhi, India

Jan 2019 - May 2019

Undergraduate Researcher

- **Generating Audio samples from Images:**

Advised by - Dr. Timothy Scott Moyers Jr

- * Developed a pipeline for bridging Audio and Visual Engines with real-time modifications
- * Learnt Non-linear mappings between projected images and audio samples
- * Developed linear and non-linear envelopes on data filters along with other common features present in DAWs

Keywords: *Computer Music, Machine Learning, Creative AI, MIDI*

WORK EXPERIENCE

- **Goldman Sachs**

Bangalore, India

Summer Analyst

May 2019 - Present

- **Regulatory Operations Engineering Team::**

- * Working on the Data Inference and Analytics (DIA) module

Keywords: *Apache Spark, Hadoop, MLlib*

- **dunnhumby**

New Delhi, India

Associate Analyst Intern

May 2017 - July 2017

- **North America Data Team::**

- * Studied the various aspects and applications of Data Science in the Business Dimension.
- * Worked on creating a large-scale implementation demo of MLlib in Apache Spark to predict user purchase patterns for various clients of dunnhumby focusing on those based in North America
- * Studied the Automation of ETL processes using Apache Airflow on Google Cloud Platform and worked in transferring data from SAS Archives to Hadoop Datalake.

* **Keywords:** *Data Analytics, Apache Spark, Hadoop*

- **Rails Girls Summer of Code**

New Delhi, India

Student Scholar

May 2017 - July 2017

- **Worked on Tessel - an IoT and Robotics development platform:**

- * Part of one of the selected 16 sponsored teams out of 192 teams worldwide.
- * Developed tutorials - Prepared Documentation, code snippets, Fritzing diagrams and a working demonstration.
- * Developed a model for a Humanoid Arm Project - Understood the idea behind product design and development.
- * Studied and worked on Reach - an ESP32 (low power BLE module), that is supposed to be Tessel’s new launch in the domain of Low power boards.

Keywords: *Open Source, IoT, Project Design*

AWARDS

- **Snap Research Scholarship 2019:** Awarded for research done in the field of Data Mining and Machine Learning. Award includes 10,000 USD and an offer to intern at Snap Research, USA.
- **Microsoft Research India Travel Grant:** Awarded travel support of 50000 INR for visiting WSDM 2019
- **ACM-W Scholarship:** Awarded travel support of 1200 USD for visiting WSDM 2019
- **Google Women Techmakers Scholarship, 2018:** Awarded to students who work for diversity and inclusion in the field of Computer Science.
- **Best Technical Poster Runner-up at GHCI 2018:** Received for the project, “Generating Clues for Gender based Occupation De-biasing in Text”
- **Dean’s Award for Innovation R&D:** Awarded to students who work on Research projects beyond coursework. Awarded for the academic years 2016-17 and 2017-18.
- **Dean’s List of Academic Affairs:** Awarded to students who demonstrate excellence in an academic year. Awarded for the academic year 2016-17.
- **Grace Hoppers Celebration India (GHCI) Scholarship, 2018:** Awarded travel grant and scholarship to attend the GHCI conference.

PROJECTS

- **img2 \LaTeX [demo]:**
 - * An end-to-end model for converting handwritten mathematical expressions to compilable \LaTeX code.
 - * Constructed a pipeline consisting of image segmentation, supervised classifiers such as CNNs, SVM, etc. and heuristics for code formation
 - * Developed as a course project for Machine Learning (CSE343) at IIIT Delhi in Monsoon 2018
 - * **Keywords:** *Machine Learning, Deep Learning, Image Processing*
- **SemEval19 Task 3: EmoContext:**
 - * Worked on the task of emotion detection in dialogue. This was a shared task for a workshop at ACL 2019.
 - * Used techniques like Data augmentation, emoji context evaluation using DeepMoji and Deep Learning models like LSTMs and ELMo.
 - * **Keywords:** *NLP, Deep Learning, Word Embeddings*
- **iDabba [code][poster]:**
 - * This project aims to improve the food storage and large scale handling of food silos in granaies
 - * Uses Computer Vision techniques with SIFT, Haar classifier and Microsoft Vision API in Python and checks the environmental conditions around it using Humidity, Temperature and Weight Sensors.
 - * Was amongst the top 10 projects in the first-year batch and received a mention in the Director’s blog [link].
 - * **Keywords:** *Computer Vision, IoT, Flask*
- **AirBnb Chatbot:**
 - * A flask chatbot developed from a curated Ontology of the AirBnb dataset
 - * Developed the Ontology from scratch - incorporating concepts like validation, SparQL querying, and N-Triple formation
 - * Incorporated human language queries using Rasa NLU API. Developed as a course project for Semantic Web (CSE632) at IIIT Delhi in Winter 2019.
 - * **Keywords:** *Flask, Ontology, Semantic Web, NLP, Chatbot*

CO-CURRICULAR ACTIVITIES

- **Talks**
 - * *Intro to Computer Vision:* Introductory talk on the self-taught topic, demonstrating basics of Images, colour, processing and basic CV algorithms. Delivered at WWCode Delhi. [slides]
 - * *Intro to Mathematics of ML:* Introductory talk on the self-taught topic, demonstrating basic ML algorithms from a mathematical point of view. Delivered at WWCode Delhi. [GitHub]
 - * *Intro to Web connectivity with Tessel:* An introductory talk about Tessel development board and its features. Delivered at LinuxChix India. [coverage]
- **Volunteering**
 - * *Summer School, IIITD:* Summer School is where middle-school children from government schools are taught basic life skills and beyond the subjects taught in school. I was a part of the Personality Development and Communication Skills team, where I taught students the importance of physical gestures like posture, speaking and listening skills and basics of english. I was also involved in the cultural development of students in Music and Dance. [coverage]