
EDUCATION

- **University of Southern California** Los Angeles, CA
Doctor of Philosophy, Computer Science Aug. 2021 (To Begin) –
- **Indraprastha Institute of Information Technology** New Delhi, India
Bachelor of Technology, Computer Science and Engineering; CGPA: 9.30/10 Aug. 2016 – Dec 2020
- **Sanskriti School** New Delhi, India
High School; Percentage: 93.8% 2014 – 2016

SKILLS

- **Relevant Courses ([†]=Graduate Level Courses):** Deep Learning[†], Machine Learning, Natural Language Processing[†], Speech Recognition and Understanding[†], Affective Computing[†], Statistical Computation[†], Semantic Web[†], Linear Optimization, Linear Algebra, Data Mining[†], Probability and Statistics, Real Analysis
- **Languages:** Python, Java, C, Bash
- **Tools & Technologies:** Git, PyTorch, Keras, MATLAB, scikit-learn, nltk, spaCy, numpy, Spark, SparQL

PUBLICATIONS

- **Brihi Joshi**, Neil Shah, Francesco Barbieri and Leonardo Neves. The Devil is in the Details: Evaluating Limitations of Transformer-based Methods for Granular Tasks. In *The 28th ACM International Conference on Computational Linguistics (COLING 2020)*.
- **Brihi Joshi***, Aditya Chetan*, 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*)
- Udit Arora, Hridoy Sankar Dutta, **Brihi Joshi***, Aditya Chetan*, Tanmoy Chakraborty. Analyzing and Detecting Collusive Users Involved in Blackmarket Retweeting Activities. In *ACM Transactions on Intelligent Systems and Technology (TIST)*. (Impact Factor: **3.971**)
- **Brihi Joshi***, Amogh Gulati*, Chirag Jain*, Jainendra Shukla. It's Not What They Play, It's What You Hear: Understanding Perceived vs. Induced Emotions in Hindustani Classical Music. In *22nd ACM International Conference on Multimodal Interaction, Late Breaking Reports (ICMI 2020)*.
- **Brihi Joshi***, Shravika Mittal, Aditya Chetan. Did You “Read” the Next Episode? Using Textual Cues for Predicting Podcast Popularity. In *First Workshop on NLP for Music and Audio (NLP4MusA) at International Society for Music Information Retrieval Conference (ISMIR 2020)*.
- Hridoy Sankar Dutta, **Brihi Joshi***, Aditya Chetan*, 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%)
- Nishtha Madaan, Gautam Singh, Sameep Mehta, Aditya Chetan*, **Brihi Joshi***. Generating Clues for Gender based Occupation De-biasing in Text [arXiv:1804.03839](https://arxiv.org/abs/1804.03839) [cs.CL]

* Equal Contribution

RESEARCH EXPERIENCE

- **Snap Research** Los Angeles, CA, USA
Research Intern Sep 2019 - Dec 2019 (Continued remotely)
 - **Understanding Granular-level Similarity of documents:**
Advised by - Leonardo Neves, Neil Shah and Francesco Barbieri
 - * Curated a dataset of news articles, with ground truth annotations of article pairs that report the same event.
 - * Benchmarked contextual embeddings generated from Transformer-based methods against simpler methods like TF-IDF for granular-level textual similarity. Formed the basis of the **Happening Now** feature in Snapchat.
 - * Currently working on understanding task granularity in cross-lingual news data.

Keywords: *Machine Learning, NLP, Representation Learning, Low-Resource Languages*
- **Laboratory for Computational Social Systems** New Delhi, India
Undergraduate Researcher Jan 2018 - Present
 - **Understanding adversarial collusive activities in OSNs [Bachelor's Thesis]:**
Advised by - Dr. Tanmoy Chakraborty
 - * Worked on detecting 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: *Unsupervised Learning, OSNs*
- **IBM India Research Laboratory** New Delhi, India
Undergraduate Researcher Nov 2017 - Sept 2018
 - **Occupational Debiasing [code][preprint]:**
Advised by - Dr. Sameep Mehta
 - * Developed a system for detecting occupational gender bias in text, by demography and time period.
 - * Developed and curated a dataset comprising of various occupation names and occupational evidences.
 - * Proposed a pipeline that detects potential gender bias in occupations

Keywords: *NLP, Machine Learning*
- **Interdisciplinary Lab for Interactive AV Development (ILIAD)** New Delhi, India
Undergraduate Researcher Jan 2019 - May 2019
 - **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.
 - * Developed a model to learn non-linear mappings between projected images and audio pieces.
 - * 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** Remote
Analyst (Full-time) Dec 2020 - Present
 - **Regulatory Operations Engineering Team::**
 - * Working on the Analytics platform for Regularity Core Engineering, which includes developing clustering scripts for downstream anomalous ticket grouping.

Keywords: *Apache Spark, Clustering*
- **Goldman Sachs** Bangalore, India
Summer Analyst May 2019 - July 2019
 - **Regulatory Operations Engineering Team::**
 - * Working on the Data Intelligence and Analytics (DIA) module

- * Responsible for intelligent join of large amount of data from multiple sources, to be used for downstream clustering tasks

Keywords: *Apache Spark, Hadoop, MLlib*

dunnhumby

Associate Analyst Intern

New Delhi, India
May 2017 - July 2017

o **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

Student Scholar

New Delhi, India
May 2017 - July 2017

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

- * Part of one of the selected 16 sponsored teams out of 190 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*

TEACHING EXPERIENCE

CSE343: Machine Learning

Teaching Assistant for a class of 150 senior undergraduate students

IIIT, Delhi
Aug 2020 - Dec 2020

CSE632: Semantic Web

Teaching Assistant for a class of 170 senior undergraduate and graduate students

IIIT, Delhi
Jan 2020 - May 2020

MTH201: Probability and Statistics

Teaching Assistant for a freshman class of 300 students

IIIT, Delhi
Jan 2019 - May 2019

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. Only scholar from India!
- **AAAI 2020 Undergraduate Consortium:** Accepted to present my thesis at the AAAI 2020 Undergraduate consortium. Includes scholarship to attend to attend AAAI 2020
- **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 and 2018-19.
- **Grace Hoppers Celebration India (GHCI) Scholarship, 2018:** Awarded travel grant and scholarship to attend the GHCI conference.

PROJECTS

- **Deep Multitask Piano Transcription** [\[report\]](#)[\[demo\]](#):
 - Implemented the Onsets and Frames baseline by [Hawthorne et. al](#) for polyphonic piano transcription.
 - Developed a smaller, efficient model, reaching the baseline performance in less time and with lesser parameters.
 - Developed as a course project for Deep Learning (CSE641) at IIIT Delhi in Winter 2020.
 - **Keywords:** *Deep Learning, Music Information Retrieval*
- **Emotional Speech to Text** [\[code\]](#):
 - Explored HMM and DL based methods to generate Emotional speech from text, along with system demonstrations.
 - Worked on developing fine-tuning strategies for SOTA methods in Speech like Tacotron and DC-TTS, to elicit emotional speech, using a low-resource dataset (EmovDB).
 - Developed as a course project for Speech Recognition and Understanding (CSE5SRU) at IIIT Delhi in Winter 2020.
 - **Keywords:** *Deep Learning, Speech Generation*
- **img2L^AT_EX**[\[demo\]](#):
 - Developed an end-to-end model for converting handwritten mathematical expressions to compilable L^AT_EX 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.
 - Designed architectures based on LSTMs and ConvNets, and explored contextual embeddings like ELMo.
 - Used techniques like Data augmentation, emoji context evaluation using DeepMoji.
 - **Keywords:** *NLP, Deep Learning, Word Embeddings*
- **iDabba** [\[code\]](#)[\[poster\]](#):
 - Built a prototype aimed at improving food storage and large scale handling of food silos in granaries.
 - 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 - incorporated 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*

LEADERSHIP

- **Women Who Code**

New Delhi, India
Aug 2017 - Present

 - **Working towards inclusion and diversity in tech:**
 - * Establishing relations with corporates to collaborate with WWCode Delhi.
 - * Responsible for mentoring new volunteers, their on-boardings, and allotting responsibilities. [\[testimonial1\]](#)[\[testimonial2\]](#)
 - * Attending meet-ups, conferences and events to expand the WWCode Delhi network and collaborating with network members for organising events and speaking opportunities.
 - * Managing the social media activities like managing the Facebook, Twitter and Meetup pages of the chapter.

- **Student Senate**
Batch Representative

- **Academic representative for the Batch of 2020:**

- * Established communication between batch students and the academic body. This includes conveying academic policies like plagiarism policies and course selection criteria with the students.
- * Handled student grievances related to academics and building their solutions.
- * Attended Undergraduate board meetings with other student representatives, Undergraduate Council Chair and Dean of Academic Affairs to discuss problems and their solutions.

CO-CURRICULAR ACTIVITIES

- **Talks**

- *Practically Machine Learning*: Taught 100 students and industry professionals basics of machine learning [\[link\]](#)
- *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*: Took sessions on Personality Development and Communication Skills for middle-school children from government schools. [\[coverage\]](#)