Sunyam Bagga

email: sunyambagga@gmail.com
website: https://sunyambagga.com/

EDUCATION

McGill University – Montreal, Canada   2017 - 2020
M.Sc., Computer Science, GPA: 4.0/4.0

Delhi Technological University – Delhi, India   2013 - 2017
B.Tech, Computer Engineering, First Class with Distinction

WORK EXPERIENCE

.txtLAB, *Research Consultant*   Feb 2020 - Oct 2021
- Built a million-page historical prose dataset with enriched NLP features from the HathiTrust Digital Library
  [published in Journal of Open Humanities Data]
- Developed computational methods for measuring the degree of 'narrativity' in 335,000 text-passages
  [published at Computational Humanities Research Conference 2021]
- Studied the effects of bias in training data for literary text classification
  [published at LaTeCH-CLfL workshop at COLING 2020]

Network Dynamics Lab, *Graduate Research Assistant*   Oct 2017 - Dec 2019
- M.Sc. thesis: detecting subtle abuse on social media using machine learning & NLP techniques
- Other projects: quantifying linguistic accommodation on Reddit; studying political polarization on Reddit

McGill University, *Teaching Assistant*   Sept 2018 - Dec 2019
- Teaching Assistant for a graduate-level course 'Natural Language Processing' in Fall 2018 & Fall 2019
- Duties: designing & grading exams, giving tutorial lectures, and grading assignments & final projects

Obsessory Pvt. Ltd., *Machine Learning Intern*   June 2016 - Aug 2016
- Implemented an Image Search Engine for a large dataset of fashion images using OpenCV

PROJECTS

Detecting offensive language on Reddit
- Identified a form of abuse inspired from the Impoliteness literature
- Annotated data through large-scale crowdsourcing using Mechanical Turk API
- Implemented state-of-the-art NLP and deep learning techniques
- Published as first-author in EACL 2021

Measuring the effects of bias in training data for text classification
- Assessed three forms of stylistic and social bias: genre, dialogue, and authorial gender
- Experimented with data augmentation techniques for NLP
- Implemented several learning models: traditional ML, deep learning, and transformers-based

Best Answer Prediction on Stack Overflow
- Devised hand-engineered features to capture different aspects of StackOverflow questions and answers
- Used word embeddings and machine learning algorithms to classify 'best' answer for a given question

TECHNICAL SKILLS

Languages: Python, SQL   **Frameworks**: transformers, sklearn, pandas, pytorch, allennlp, nltk, gensim, tweepy

OTHER ACHIEVEMENTS
- Scored 332/340 in the Graduate Record Examinations (GRE) - Quantitative Section: 170/170, 2016
- All India Rank 4,676 in JEE Mains - 1.4 million students participated, 2013