208 Williams St
Ithaca, NY 14850
(+1) 571-831-1162

Kabir Walia

kabirwalia8300.github.io

github.com/kabirwalia8300
linkedin.com/in/kabirwalia
ksw63@cornell.edu

Education

Cornell University – GPA: 3.82/4.00 (Dean's List) Aug 2018 – Dec 2021, Ithaca, NY
College of Engineering – B.S Computer Science with a minor in Operations Research and Management Science

Relevant Coursework – Natural Language Processing, Machine Learning for Intelligent Systems, Computer Vision, Advanced Language Technologies, Algorithms, Principles of Large-Scale ML, Big Messy Data, Operating Systems, Probability and Statistics

Professional Experience

Microsoft – Data and Applied Scientist Intern Nov 2020 – Present, Bellevue, WA
Dynamics 365 Fraud Protection (DFP) team

- Joining the DFP team at Microsoft as a Data and Applied Scientist Intern for Summer 2021

Amazon – Research Engineer Intern Jun 2020 – Sept 2020, Bangalore, India
Amazon India Machine Learning team

- Owned development and integration of an automated data labeling workflow (Active Learning) into an internal AutoML tool to streamline business workflows for 50+ non-technical teams globally
- Decreased costs of labeling data for text-classification tasks by over 65% while maintaining 0.95 precision and recall metrics
- Benchmarked BALD (Bayesian Active Learning by Disagreement) against classical uncertainty and diversity–based sampling methods on 8 Amazon Business datasets – laying the foundation for further internal research on neural active learning
- Built an interactive UI for facilitating the labeling task using Streamlit, AWS Sagemaker, Ground Truth, Step Functions, S3

Cornell NLP Group – Undergraduate Researcher Sept 2019 – Present, Ithaca, NY
Member of Prof. Claire Cardie's Natural Language Processing lab

- Exploring neural architectures and hierarchical embeddings for few-shot and low-resource Named Entity Recognition (NER)
- Developing a novel architecture leveraging the recurrence mechanism of Transformer-XL and conditional language modeling system of the CTRL model to emulate argumentation strategies like card-stacking for producing higher quality counterarguments

Mu Sigma – Software Engineering Intern Jun 2019 – Aug 2019, Austin, TX
India's largest data-driven management consulting firm

- Led a team of 3 interns to setup the 1st US-based AI interactive lab to showcase technological capabilities during client demos
- Implemented anomaly detection algorithms like recursive residuals, achieving 71% accuracy on a client's simulated IoT data
- Developed an automated office assistant, *muNerva*, using Flask, Alexa Voice Services, Redis, and OpenCV

Leadership

Cornell Cup Robotics – CS Team Lead Feb 2019 – May 2021, Ithaca, NY
Oversaw development of a distributed system to control Minibot, an educational robot to teach children to code using Google Blockly

- Collaborated with Systems, Electrical and Mechanical sub-team leads to define workflows and implement objectives for Minibot
- Managed an 18-member software team to complete computer vision, database management, and computer networking projects
- Optimized the single-camera vision system by decreasing sensor to GUI lag from ~35s to ~5s by restructuring code base
- Regulated public GitHub, led scrums and code reviews, and pitched marquee features to C–suite of partnering ed-tech startup

Cornell University – Teaching Assistant and AEW Instructor Aug 2019 – May 2021, Ithaca, NY
CS 4670 (Computer Vision), CS 4786/5786 (Machine Learning for Data Science) & MATH 2940 (Linear Algebra)

- Selected as 1 of 2 sophomore teaching assistants for a graduate level machine learning course with 180+ students
- Guided office hours, created and graded assignments, and handled questions on Piazza to assist students' learning experience
- Lectured on linear algebra concepts in a weekly workshop to a class of 18 students (on campus and on Zoom)

Projects

TriBlank – Multiple Entity Blanking for Relation Learning [GitHub] Feb 2021 – May 2021, Ithaca, NY

- Designed neural network architectures for semantic relation extraction trained by blanking out entities in text
- Validated method on DocRED dataset, supporting hypothesis about the impact of contextual entities on semantic relations between target entities

TextDirect – SMS-based routing [GitHub] Dec 2019 – Jan 2020, Ithaca, NY

- Launched a navigation service providing step by step Google Maps directions as text messages (Twilio, Google Cloud Platform)
- Aimed at benefitting feature phone users and users in areas where internet/mobile data penetration is limited

Skills & Activities

- **Software:** Python | Java | C | Scala | PyTorch | TensorFlow | AWS | SQL | Databricks | Hadoop | Docker | Git | Bash | Linux
- **General:** Leadership | Design Thinking | Project Management | Agile | Scrum | Resource Planning | Teaching
- **Activities:** Theta Tau Professional Engineering Fraternity | Cornell European Business Society | Intramural Soccer
- **Interests:** Linguistics | Tennis | Acting | War Documentaries | Hiking | Harmonica | Cooking | Manchester United