SAYALI GHODEKAR

New York, NY 10027

+1(929)677-0630◇ smg2280@columbia.edu ◇ LinkedIn ◇ GitHub ◇ Google Scholar

EDUCATION

Columbia University *New York, NY*
Master of Science, Computer Science *Expected Dec 2022*
Relevant Coursework: NLP, Fundamentals of Speech Recognition, Adv Spoken language processing, Deep learning for Computer Vision, AI, ML and Climate.

Savitribai Phule Pune University *Pune, India*
Bachelor of Computer Engineering, GPA: **3.97**/4 *June 2019*
Relevant Coursework: Machine Learning, Data Analytics, AI and Robotics, Soft Computing and Optimization, Engineering Mathematics, Cloud Computing, High Performance Computing.

WORK EXPERIENCE

RingCentral Inc *Mumbai, India*
Data Scientist (Conversation AI Team) *Dec 2020 - July 2021*
- Created innovative Machine learning strategies for NLP problems, leading to integration of AI-based services in RingCentral's video communication platform for 300K businesses worldwide.
- Built, deployed and scaled abstractive summarization service to extract key insights from conversations, decreasing system latency. Implemented data collection, models and client libraries, dockerization, and Kubernetes deployments.
- Improved qualitative performance of post-meeting tools through text-generation services including Coreference-Resolver, question-generator, text-paraphraser, and multilingual summarizer.

DeepAffects - Conversation-Intelligence Startup *Mumbai, India*
Data Scientist *Jan 2020 - Dec 2020*
- Developed and scaled DeepAffect's abstractive summarization API, leading to a 10% increase in model performance. Collaborated with quality team for continuous testing, processing over 1M+ minutes of audio into summaries.
- Enhanced DeepAffect's conversation metrics stack through dialogue act tagging, intent classification, question and answering systems, reducing time to market these APIs by 75%.
- Proposed and implemented a topic segmentation algorithm for multi-turn dialogue data, demonstrating an improvement in model performance across all NLP pipelines.

Center for Indian Language Technologies, IIT Bombay *Mumbai, India*
Research Intern under the guidance of Prof. Pushpak Bhattacharyya *June 2019 - Dec 2019*
- Designed experiments for Cognate Detection with Siamese neural networks, evaluating word embedding models and beating baseline by 71%.
- Collaborated with Cognitive NLP group to derive gaze features for cognate identification. Reduced annotation cost by leveraging predictive models, demonstrating a 12% improvement over baseline.
- Published and presented cognate study findings in ACM CoDS-COMAD 2021 [link] and EACL 2021 [link]. Awarded **best paper honorable mention at EACL 2021**.
- Conceptualized and designed novel Textual History Analysis Tool to capture historical evolution of texts using computational phylogenetics. Managed tool deployment 2 weeks prior to deadline, resolved technical conflict during presentation.

TEACHING

Natural language processing *New York, NY*
Teaching Assistant for Graduate-level course at Columbia Computer Science *Fall 2021*

ACADEMIC PROJECTS

Code-Switched Automatic Speech Recognition System *New York, NY*
- Built an ASR system for recognizing Code-switched text for the low-resource Hindi-English language pair using the Kaldi toolkit. Created my self-recorded dataset and achived a WER of 36 on the test set.

TECHNICAL SKILLS

Python, R, C++, Torch, Tensorflow, Keras, Kaldi, scikit-learn, OpenCV, Docker, Kubernetes, AWS, Javascript, PHP, MySQL, MongoDB, PowerShell, Basics of Google Cloud Services.