Aneri Rana

arana@umass.edu | (718)808-5622| linkedin.com/in/aneri-rana-09bb87137 | github.com/anerirana

EDUCATION

University of Massachusetts Amherst, USA Exp. Graduation **May 2024**
Expected to graduate with a Master of Science in Computer Science
Courses: Advanced Machine Learning Theory, Natural Language Processing, Applied Statistics **3.856 GPA**

TECHNICAL SKILLS

- Programming Languages: Python, R, MATLAB, BigQuery, SQL, C++, Java, Terraform, Ansible, Groovy, HTML.
- Tools and Technologies: PyTorch, TensorFlow, HuggingFace, Keras, Data Studio, Tableau, Apache Ariflow, Apache Spark, Google Cloud Platform.

PROFESSIONAL EXPERIENCE

C3.ai, Redwood City, California
Data Science Intern **June 2023 – Aug 2023**

- Developed a robust logs error analysis framework for C3's core platform, reducing service interruptions by 30% and enhancing overall user experience.
- Identified and prioritized 30 critical bug fixes in the platform codebase through efficient analysis of error log summaries, utilizing regular expressions and the Drain3 clustering algorithm.
- Established an efficient pipeline integrating server metadata into error logs summary model for root cause analysis of server shutdowns/restarts, reducing bug discovery and debugging time from 1 day to 30 mins.
- Designed and documented a self-help Jupyter Notebook leveraging the error summary model empowering users to independently troubleshoot and resolve blockers, optimizing resources for the platform team.

HSBC Technologies, Pune, India
Senior Software Engineer | *Data Science* **March 2020 – Aug 2022**

- Developed a **multi-task deep learning** model with **0.93** precision and **0.96** recall, for automatic repair of transaction fields, reducing manual processing of over 200k failed transactions/month in the UK.
- Coordinated a team of 5 Data Scientists, Data Engineers and Business Analysts to engineer a scalable solution for **long-term deployment and monitoring** of the Automatic Repair ML model across multiple countries.
- Trained a **BiLSTM** model achieving an F1-Score of **0.94** for address classification in payments, replacing manual efforts and generating audit reports for HSBC through confidence-scored predictions in Data Studio.
- Implemented a GCP solution for batch training and prediction of Machine Learning models in production scheduled using **Apache Airflow** on **Google Kubernetes Engine**.
- Built a distributed cloud computing platform for parallel processing of Machine Learning models on Apache **Hadoop YARN** Clusters with **Apache Spark**.
- Orchestrated the development of a plug-and-play **API** infrastructure on Google Kubernetes Engine, facilitating the **deployment** of the first two **ML models** with self-help documentation for seamless integration into production.

Software Engineer | *Mobile Application Development* **Aug 2018 – Feb 2020**

- Devised HSBC Kinetic iOS app using Swift; Designed the Open Banking, Cards, Transaction, and Login session management modules of the application.
- Designated as iOS Champion for performing code reviews, conversing with business stakeholders to understand requirements and propose feasible solutions, and supporting DevOps tools.

PROJECTS, PATENT, AND PAPER

Common Sense Reasoning through Winograd Schema Challenge
University of Massachusetts Amherst, Natural Language Processing Project, Jan 2023

- Conducted in-depth research and analysis of state-of-the-art large language models for common sense reasoning tasks, specifically focusing on the Winograd Schema Challenge (WSC).
- Formulated the WSC problem as a text-generation task and conducted experiments with **zero-shot/few-shot prompting**, **Chain of Thought** (CoT), and **parameter efficient prefix tuning** on **GPT-3.5**, **Flan-T5** in small, base, large, XL, and XXL model versions.
- Employed cosine similarity and random sampling techniques to strategically select in-context examples for few-shot and CoT prompting.

- Developed a novel heuristic called "top-5" to identify and utilize five sentences with the highest frequency from the DPR train data set, each associated with a specific thought, as input for the CoT prompt.
- Attained a remarkable **91.22%** accuracy using CoT prompting with the top-5 heuristic on the **Flan-T5-XXL** model (11B parameters), demonstrating comparable performance to the current state-of-the-art accuracy of 91.28% on the Allen Institute Leaderboard.

Mars Spectrometry | Detect Evidence for Past Habitability

ML Challenge, NASA, June 2022

- Evolved gas analysis mass spectrometry data from Mars exploration missions by curiosity rover was given by NASA to predict chemical compositions which indicate past livable conditions on Mars.
- Ensembled various combinations of **metric learning** algorithms like LMNN, NCA, ITML and classifiers such as **Extra Trees** and **LightGBM** to predict composition of chemical compounds in the geological samples.
- Attained **18th position globally** among 713 participants with a log loss of 0.14 on the test data set. https://www.drivendata.org/competitions/93/nasa-mars-spectrometry/leaderboard/

System and Method for Classification of Hate Speech

- Engineered and patented an innovative system for the classification of audio/video content on social media as hate speech. Patent pending, **Indian Patent Application 201921052170.**
- Collected and annotated 1k video dataset from YouTube and Twitter to train a **multimodal deep learning** model that combines emotional features of speech with semantic features of the text.
- Overcame the challenge of a limited dataset by leveraging **transfer learning** from pre-trained text and emotion models.
- Extracted text features from the last layer of a **BERT** model fine-tuned using PyTorch on hate speech Tweets.
- Extracted emotion features from a **CNN** model pre-trained on an existing audio dataset to predict emotion attributes, i.e., valence, arousal, and dominance.
- Authored a paper on the same, currently under review by the ACM Transactions on Intelligent Systems and Technology. https://arxiv.org/abs/2202.06218

Game AI

University of Massachusetts Amherst, Reinforcement Learning Project, Nov 2023

- Applied **Actor-Critic** and **Reinforce with Baseline** Reinforcement Learning (RL) algorithms to train AI agents proficient in playing OpenAI Gymnasium's cart-pole, mountain car, and cliff walking games.
- Developed custom gaming environments, including 687-GridWorld and Snake game, for specialized RL training of AI agents.

COURSES AND CERTIFICATIONS

- **Computer Science Research Mentorship Program Scholar** [CSRMP 2023a] - Accepted to a mentorship program by **Google** in support of my pursuit of machine learning research pathways, Feb 2023 to Jun 2023.
- Machine Learning, by Stanford University, Coursera in Nov 2019.
- Google Cloud Certified “Associate Cloud Engineer” in Aug 2019.
- Developing applications with Google Cloud Platform specialization, by Google, Coursera in Mar 2019.

VOLUNTEERING

- Programmed new features, fixed bugs when identified, and resolved issues raised by the users for **Keras**, **Snorkel**, and Avalanche **Open Source** Projects, May 2021 – March 2022.
- Managed the website and developed new features (in python) as the Super Volunteer and Web maintainer for Women in Machine Learning (WiML) which aims to enhance the experience and impact of women in ML, Oct 2021 – Aug 2022.
- Facilitated a one-hour breakout session on “Leveraging Open-Source Tools for NLP” at the **International Conference for Machine Learning** (ICML, a top-tier ML conference) with research associates from the Georgia Institute of Technology and the University of Toronto in July 2021.
- Selected as an **emcee** to host talk and keynote sessions for an entire day at the Open Data Science Conference in Sept 2021. Introduced speakers, handled live broadcasting and managed issues during the session.