Arjun Ramesh Rao

mailarjunrao@gmail.com | arjun.fyi | arjun-rao | arjunra0 | @arjunra0

Education

University of Colorado, Boulder — *Boulder, US*

M.S. in Computer Science — *Aug. 2019 - May 2021*

- GPA: 4.0/4.0
- **Graduate Teaching Assistant** for CSCI 1300: Starting Computing (Fall 2020, Spring 2021)
- **Graduate Research Assistant** under Prof. Sidney D'Mello (Spring 2020)
- **Graduate Student Staff** for CSCI 3202: Intro to AI (Fall 2019)
- Awarded Lloyd Botway Award for Outstanding Master's students

Ramaiah Institute of Technology — *Bangalore, India*

B.E. in Information Science and Engineering — *Aug. 2014 - June 2018*

- CGPA: 9.41/10
- **Best Outgoing Student** (Batch of 2014-2018), Dept. of Information Science & Engineering

Industry Experience

Microsoft Corp. — *Redmond, US*

Software Engineer — *June. 2021 - Present*

- Working on recommendations and ranking for Microsoft News and Feeds
- Technologies Used: C#, .Net Framework, Python

Microsoft Corp. — *Redmond, US*

Software Engineer Intern — *May 2020 - Aug. 2020*

- Worked on the Core Ranker Team within Microsoft News and Feeds
- Built automated pipelines for continuously evaluating and improving Content Classification models.
- Built UHRS apps for obtaining crowd sourced training data for document classification.
- Technologies Used: C#, Python, .Net Framework, JavaScript, UHRS

Stride.ai Inc. — *Bangalore, India*

NLP Engineer — ***July 2018 - May 2019***

- Built pipelines to automatically train machine learning models for document classification and information extraction.
- Built a **custom python library** that allows users to train, evaluate and perform inference on document classification tasks with limited labeled data.
- Contributed to the Django backend pipeline for model training, evaluation and inference.
- Worked on information extraction use-cases for **identifying key datapoints from documents** using NLP based models and custom heuristic rules.

Stride.ai Inc. — *Bangalore, India*

Research and Development Intern — *Jan.- Apr. 2018, Nov. - Dec. 2016*

- Worked on **optimizing Named Entity Recognition (NER) models** in a production project to improve performance by combining standard NER datasets with context specific data.
- Built **scrapers** for automatically extracting information from multiple websites involving **simulation of complex UI interactions**.
- Technologies Used: Python, Tensorflow, Keras, Gensim, Selenium

Google Inc. — *San Francisco, US*

Developer Programs Engineer Intern — *June 2017 - Aug. 2017*

- Built tools to help **track code repositories** and generate **consolidated notifications** for events like issues, comments, etc., for faster triage of issues.
- Part of the project was **released as open source software**, and can be found at github.com/GoogleCloudPlatform/issuetracker
- Technologies Used: Google Cloud Datastore, Google BigQuery, Google App Engine, GoLang, Angular

Academic Experience

University of Colorado, Boulder
Boulder, US

Emotive Computing Lab, Institute of Cognitive Science
Jan. 2020 - May 2021

- Worked under Prof. Sidney D'Mello at the Emotive Computing Lab.
- Contributed to active research on modeling collaborative problem solving processes and discourse using state of the art natural language processing techniques and multi-modal machine learning.
- Worked on modeling bias in machine learning models for apparent personality prediction in one way behavioral interviews.
- Technologies Used: **Python**, **PyTorch**, **AWS**

Ramaiah Institute of Technology
Bangalore, India

Senior Project - The Milo IDE (miloide.github.io)
Sept. 2017 - April 2018

- Built a web-based IDE to help students with no prior programming experience learn Machine Learning and Linear Algebra.
- Customized **Google's blockly project**, and designed a visual programming language that supports data science operations.
- Implemented a data explorer with built-in datasets along with support for using custom numeric, image and textual datasets.
- Implemented common ML algorithms using **Tensorflow.js** as blocks and used **D3.js and Plotly.js** for interactive visualizations.
- Presented and published a paper based on a user study with the IDE at IEEE VLHCC 2018 (See Publications).
- Technologies Used: **Node**, **Javascript**

Select Publications

Say What? Automatic Modeling of Collaborative Problem Solving Skills from Student Speech in the Wild
(Virtual) Paris, France

Samuel L Pugh, Shree Krishna Subburaj, **Arjun Ramesh Rao**, Angela EB Stewart, Jessica Andrews-Todd, Sidney K D'Mello
June. 2021

- Proceedings of the Educational Data Mining Conference 2021 [PDF]
- We investigated the feasibility of using automatic speech recognition (ASR) and natural language processing (NLP) to classify collaborative problem solving (CPS) skills from recorded speech in noisy environments.

Multimodal, Multiparty Modeling of Collaborative Problem Solving Performance
(Virtual) Utrecht, Netherlands

Shree Krishna Subburaj, Angela EB Stewart, **Arjun Ramesh Rao**, Sidney K D'Mello
Oct. 2020

- Proceedings of the 2020 International Conference on Multimodal Interaction, pp. 423-432. [PDF]
- Analyzed data from 101 triads engaged in computer-mediated collaborative problem solving (CPS) in an educational physics game.
- Investigated the accuracy of machine-learned models trained on facial expressions, acoustic-prosodics, eye gaze, and task context information, computed one-minute prior to the end of a game level, at predicting success at solving that level.
- DOI: 10.1145/3382507.3418877

Milo: A visual programming environment for Data Science Education
Lisbon, Portugal

Arjun R Rao, Ayush Bihani, Mydhili K Nair
Oct. 2018

- Proceedings of 2018 IEEE Symposium on Visual Languages and Human-Centric Computing **(VL/HCC'18)**, pp. 211-215. [PDF]
- Designed and implemented a novel **visual programming environment** to help novice students and non-programmers learn **Data Science and ML concepts** using block based programming.
- DOI: 10.1109/VLHCC.2018.8506504

Context Based Approach for Second Language Acquisition
New Orleans, USA

Nihal V Nayak, **Arjun R Rao**
June 2018

- System paper for Duolingo's shared task on Second Language Acquisition Modelling **(SLAM 2018)**. [PDF]
- Published in the Proceedings of the **NAACL-HLT Workshop** on Innovative Use of NLP for Building Educational Applications **(BEA at NAACL 2018)**.
- Trained a logistic regression model to predict the likelihood of a student making a mistake while answering an exercise on Duolingo. Made use of features inspired by research in **code-mixed language learning** where context plays an important role.
- Result: **AUROC scores for English/Spanish = 0.821**, Spanish/English = 0.790 and French/English = 0.812. **2nd best linear model**, finished 9th overall in SLAM 2018

Skills

Programming Languages	**Python**, **C# TypeScript**, **Go**, **Java**, C/C++, PHP
Platforms & Frameworks	**.Net Framework**, **PyTorch**, **Tensorflow**, **Angular**, **Cloud Development**, **DJango**, **Android**
Languages	**English**, Hindi, Kannada