Rajath Rao

+1 408-300-4564 • rajath.rao@stonybrook.edu • Website • LinkedIn • US Citizen

EDUCATION

M.S. Data Science Stony Brook University, *New York* *Aug 2023 - Present*

B.S. Computer Science & Engineering University of California - Irvine, *California* *Sep 2019 - Mar 2023*

- *Dean's Honor List 2022*

PROFESSIONAL EXPERIENCE

NLP Research, HLAB at Stony Brook University, *New York* Aug *2023 - Present*

- Determined two novel audio features for predicting PTSD symptom severity from 9/11 World Trade Center victims/first-responders
- Leveraged *'OpenWillis'* and *'OpenAI Whisper'* to extract speech characteristics from audio to create a *differential language analysis feature table*
- Analyzed duration and frequency of extremes using statistical outlier detection and with k-fold cross validation
- Training a regression model for assessing PTSD-Check-List (PCL) scores based on prosody and speech characteristics with natural language features
- **Utilized:** *Python, OpenWillis, DLATK, HuggingFace, PyTorch, NCCL, CUDA, SQL, Academic Writing*

Software Engineer Intern, Intel, *Santa Clara, CA* Jun *2021 - Apr 2023*

- Proposed and developed machine learning models for failure prevention, predictive maintenance, pattern recognition, and automation scripting with MPI for Intel's HPC
- **Hard Drive Imminent Failure Predictor, Intel**
  - Developed an ensemble of stochastic gradient descent neural networks for predicting hard drive failures up to 4 months before they occur and estimating the number of days till the imminent failure
  - Instantiated persistent MongoDB database with daily accumulation of S.M.A.R.T (Self-Monitoring, Analysis-and-Reporting-Technology) hard-drive data from 1,000+ host systems
  - Leveraged information gain using confusion matrix, correlation coefficients, hyperparameter tuning, cross-validation, and ROC/AUC/MSE scores to train and validate model accuracy
  - Trained predictor notifies of imminent failures up to 4 months in advance with a 97% accuracy on test environment servers saving terabytes of data loss alerting maintenance teams for hard drive backups
  - **Utilized:** *Python, Machine Learning, Feature Engineering, Jupyter, Sci-Kit Learn, Flask, MongoDB, Docker*

SKILLS

Languages: *Python, C/C++/C#, Java, JavaScript, TypeScript, HTML, CSS, Go, R, SQL, LISP, Prolog, Perl, Dart, Verilog, Linux*
Skills: *Computer Vision, NLP, AI/ML, PyTorch, TensorFlow, CUDA GPU, OpenCV, HuggingFace, Algorithms, Data Analysis, AWS, Azure, Google Cloud, MPI, Apache Spark, Kafka, React/Node.js, REST, Docker, Kubernetes, Git, JIRA, Confluence*

PROJECTS

Reggelia - Language Learning with GenAI, *Personal Project* *Jul 2023 - Nov 2023*

- Spearheaded research in *'shallow-fusion'* Automatic Speech Recognition(ASR) for bilingual utterance transcription
- Developed and fine-tuned OpenAI's Whisper model achieving 3% word-error-rate on transcribing English-Spanish
- Applied dimensional reduction technique (PCA) on web-scraped data to create embeddings for a parallel corpus of English-Spanish verbal/textual utterances
- Leveraged OpenAI API for fine-tuning *'gpt-3.5-turbo'* model with knowledge base along with few-shot prompting
- Optimized AzureSQL database schema to efficiently index user data with normalization techniques
- **Utilized:** *Azure, OpenAI, React/Node.js, Python, HuggingFace, PyTorch, Docker, JIRA, Git*

Panoptic Image Segmentation - CV Object Classification, *Personal Project* *Dec 2022 - Feb 2023*

- Developed a U-Net model to robustly classify and segment objects in images containing varying levels of noise and camera distortion
- Extended existing panoptic segmentation architecture by coupling semantic and instance segmentation masks to train on preprocessed, noisy images
- Developed a custom noise reduction pipeline that effectively reduced noise artifacts while preserving important image features [Higher Signal-to-Noise Ratio (SNR)]
- Conducted thorough quantitative evaluation using IoU/mAP to measure model performance on training/validation

- ❖ Achieved about 15% improvement in segmentation accuracy compared to baseline models on noisy datasets
- ❖ **Utilized:** *Python, Computer Vision, Data Validation, TensorFlow, Keras, OpenCV, Jupyter*

Real-Time Location Services (RTLS) via Bluetooth (BLE), *Consulting* *Apr 2023 - Aug 2023*

- ❖ IoT solution with BLE anchors/tags for RTLS and resident-monitoring currently in use at Roseleaf Senior Care
- ❖ Leveraged AWS IoT/DynamoDB to route Received-Signal-Strength-Indicator (RSSI) packets from tags to anchors
- ❖ Developed AWS Lambda functions to triangulate locations using 3-point trilateration based on RSSI-distance values estimated with a logistic regression model
- ❖ Designed a mobile application with React Native for an organized display of resident data and real-time locations
- ❖ **Utilized:** *AWS, React/Node.js, Python, IoT, BLE, Git, Devpost*

B.S. Capstone: Autonomous IoT Shopping Cart with Intelligent Tracking, *UC - Irvine* *Oct 2022 - Mar 2023*

- ❖ Spearheaded the development of an IoT based autonomous shopping cart with user-following, lane-correction, object-collision, and product-search features
- ❖ Created the Mealy Machine (FSM) structure for the autonomous drive-state decision making process of the cart
- ❖ Devised a WiFi triangulation algorithm using 3-point trilateration of Received-Signal-Strength-Indicator (RSSI)
- ❖ **Utilized:** *Arduino, ESP8266, C/C++/C#, Python, Scikit-Learn, Firebase, MQTT Broker, Mealy Machine (FSM), GitHub*

HotSpot: COVID-19 Real-Time Alerts, *MHacks 13 - University of Michigan* *Jul 2020 - Dec 2020*

- ❖ Alerted our users with real-time cases using iOS Radar and Geolocation data pushed to Google Firebase
- ❖ Leveraged Google Cloud API to geolocate user positions and modeled imminent COVID cases with kNN script for geofencing the infected user network
- ❖ Awarded with “Wolfram Award for Top 30 Hacks,” “Best COVID-19 App (G Cloud),” and “Most Creative Radar Hack”
- ❖ **Utilized:** *Javascript, React/Node.js, Swift, Python, TensorFlow, Firebase, Radar, Google Maps API, Git, DevPost*