Diya Saha

📞 (408)-480-7445 ✉ diya.cottonian@gmail.com 🌐 portfolio in diya-saha ○ DiyadotSaha

Education

University of California, Irvine Sep. 2023 – Jun. 2025
Master of Science in Computer Science

University of California, Santa Cruz Sep. 2019 – Jun. 2023
Bachelor of Science in Computer Engineering with specialization in Robotics; Minor in CS

Experience

Canine Cancer Research with Dr.Pierre Baldi July 2024 – October 2024
Graduate Researcher *Irvine, CA*
- Utilized LLMs to generate automated summaries and overviews of the clinical reports, aiding in the automation of the clinical report generation process
- Extracted information of interest using an OCR model from unstructed medical reports of canine subjects in the Anivive Lifesciences database and used machine learning models to study patterns in detected biomarkers

Low Resource Language Research September 2024 – November 2024
Graduate Researcher *Irvine, CA*
- Fine-tuned mBART on a curated Nahuatl-Spanish corpus (government documents, religious texts, children's books) to develop a bidirectional translation system for low-resource languages.
- Engineered a parallel corpus, applied custom evaluation metrics, and used transfer learning to mitigate corpus biases and enhance translation performance.

DSights.Inc June 2022 - September 2022
Software Research Intern *Remote*
- Used Natural Language Processing Techniques to automatically compare restaurant menus from web pages
- Explored different ideas for a prototype to help our client find similarities between restaurant menus based on words and phrases and submitted a written report summarizing the findings
- Conducted analysis on the menus publicly available and explored different methods to visualize data using Python libraries like Matplotlib, Pandas and Seaborn, and Google sheets

Projects

Context-Aware Chatbot using RAG | *Langchain, FAISS, HuggingFace* June 2024
- Developed a context-aware chatbot tailored for UCI students, facilitating course inquiries and providing context-specific responses
- Executed end-to-end machine learning pipeline from creating a vector database to create a prompt template to help the chatbot answer questions
- Link to GitHub repository

Determining readmission rate on diabetic patients | *Python, Pandas, Matplotlib, Sklearn* September 2023
- Utilized a comprehensive decade-long dataset from 130 US hospitals to analyze patterns related to the early readmission of diabetic patients, employing classification, ensemble, and neural network models.
- Implemented data imputation, oversampling techniques to address class imbalance, and applied feature selection for model optimization.
- Link to GitHub repository

UAV Model Simulator, Evade and Avoid | *Python, GUI, PyQt5, Matplotlib* March 2023
- Developed and integrated evasive maneuvers functionality into the flight simulator, implementing new control laws to ensure avoidance of explosive forces along the aircraft's trajectory.
- Enhanced autopilot model by incorporating a robust state machine and refined waypoint following algorithm, resulting in visually and computationally effective aircraft maneuvers while avoiding potential explosive hazards.
- Link to GitHub repository

Technical Skills

Proficient in Programming Languages: Python, Java, Git, C/C++, HTML/CSS, JavaScript, SQL, Assembly
Developer Tools:Google Cloud Platform, Amazon Developer Tools, Vivado, Figma, Langchain, HuggingFace
Technologies/Frameworks: PyTorch, TensorFlow, Keras, Scikit-learn, GenAI, RAG, LLMs, MLOps

Honors / Certificates

- Certified Amazon Web Services *AWS* AI Practitioner
- Received Dean's Honor on multiple quarters for being in the top 15% of my class.
- Google Professional Certification for Data Analytics, Google Professional Certification for UI/UX Design