SIDHARTH BABU

 (+1) 408-649-0393   sidharthbabukish121@gmail.com   linkedin.com/in/sidharth-babu   SidharthBabu121

Education

University of California, Santa Cruz   **Sep 2022 – June 2024**
Masters of Science in Machine Learning (Natural Language Processing) - GPA: 3.7   Santa Cruz, USA

University of California, Merced   **August 2018 – May 2022**
Bachelor of Science in Computer Science and Engineering   Merced, USA

Technical Skills

Programming: Python, C/C++, Java, JavaScript, HTML/CSS, R, MATLAB, React
Frameworks: PyTorch, TensorFlow, Keras, HuggingFace, Spacy, NLTK, Pandas, OpenCV, NumPy, Matplotlib, SciPy
Developer Tools: Git, Docker, Kubernetes, Amazon Web Services (AWS), VS Code, Visual Studio, PyCharm, Eclipse

Work Experience

AI Software Engineer   **Jan 2025 – Present**
SuperIntro | *Flask, MongoDB, React, Node.js, TypeScript*   San Francisco, CA
- Intrumental in developing and optimizing scalable RESTful APIs with Flask and ExpressJS (Node.js) to enable seamless data integration and real-time data handling.
- Building responsive, and dynamic interfaces with React and TypeScript, ensuring seamless user engagement.

NLP Graduate Student Researcher   **Jan 2023 – Dec 2023**
UC Santa Cruz | *Python, NLP, LLM, pBBLEU, UniEval, Sentiment Analysis*   Santa Cruz, CA
- Developed a Natural Language Generation (NLG) pipeline across **9** knowledge-grounded dialogue acts and **7+** domains using prompt-based transfer learning and multi-hop knowledge relationships to improve conversational coherence.
- Leveraged the pipeline to generate and refine a utterance-level corpus with over **132k labeled utterances** to have a **semantic accuracy score over 77%** and a **97% dialogue act accuracy** score in order to generate a comprehensive dialogue-level corpus.
- Attained a **competitive UniEval scores (over 85%)** for the dialogue corpus; targeting coherence, naturalness, and engagingness to create interesting and engaging conversational interactions in Open-domain Dialogue Systems.

TITANS – Math & Analytics Intern   **June 2022 – August 2022**
Sandia National Labs | *Python, Pytorch, OpenCV, Kubernetes, Pandas, Seaborn, SciPy*   Albuquerque, NM
- Conducted a detailed analysis of correlations and relationships within measured variables, revealing valuable insights into electrical test data for Sandia's commercial-off-the-shelf (COTS) program.
- Utilized and improved upon exploratory data analysis techniques to compare test data from over **8 different 3rd party Sandia affiliated sites** before generating appropriate data visualizations for new use case opportunities.

UCM-LLNL Summer Data Science/ML Intern   **May 2021 – June 2021**
Lawrence Livermore National Labs | *Python, Tensorflow, Keras, Pandas, Matplotlib, Astropy*   Merced, CA
- Developed and optimized several supervised and unsupervised machine learning models with a **95.4% precision and 96.3% recall rate** on recent time-domain optical astronomy data to detect, distinguish, and characterize Near Earth Objects (NEOs).
- Delivered a **25 minute presentation to 50 LLNL personal** detailing findings and potential applications of the developed predictive models and their corresponding solutions.

UCM-JGI Summer Research Intern   **June 2020 – August 2020**
DOE Joint Genome Institute | *Python, Pytorch, AWS, SciPy*   Merced, CA
- Evaluated and enhanced a hybrid clustering strategy with the bio-python library on the Lawrencium Super Cluster, resolving the big data challenges in metagenomic assembly for a dataset of **over 30,000+ reads.**
- Contributed to the testing and implementation of a new scalable algorithm, Apache SpaRC, **achieving an 80% to 90% accurate** taxonomic profile for the de novo reconstruction of over 8 artifical BioCrust genomes.

Personal Project Highlights

Clickbait Spoiling via Spoiler Generation | *PyTorch, Python, HuggingFace, LLM*
- Implemented and strengthened several natural language processing models, achieving **a 16.3% accuracy improvement** in generating spoiler sentences to combat clickbait. Built two transformer based extractive Q&A models BERT-BASE-SQuAD-2 and RoBERTa-BASE- SQuAD-2 with **over 89% in accuracy** to improve spoiler generation.

Project Protect: HealthHost | *Javascript, React, HTML/CSS, Backend Dev, Frontend Dev*
- Developed and launched a multi-platform mobile application to improve access to healthcare for the multi-ethnic communities in Merced County with over **500+ users.**
- Implemented audio features for Hmong, English, and Spanish sections while collaborating with the Content Team to update a non-coder-friendly maintenance manual to facilitate new markets and better serve the existing client base.
- Won **top honors and a $2,500 cash prize** at the Fall 2019 Innovate to Grow Competition within Merced county.