Subharthi Saha

LinkedIn: linkedin.com/in/subharthi-saha/
Portfolio: shubhsaha.com

Email: subharth@usc.edu
Mobile: +1 (213) 667-7304

Education

- **University of Southern California** — Los Angeles, USA
  Master's of Science - Machine Learning and Data Science; GPA: 3.81/4 — *Aug 2021-May 2023*
  Courses: *Machine Learning, Deep Learning, Databases, Applied and Cloud Computing, Data Structure and Algorithms, Linear Algebra, Probability Theory, Digital Signal Processing*

- **Vellore Institute of Technology** — Vellore, India
  Bachelor of Technology - Electronics and Communication Engineering; GPA: 8.95/10 — *Jul 2017-Jul 2021*
  Courses: *Machine Learning, Data Mining and Predictive Analysis, Computer Vision, Big Data Analytics*

Technical Skills

- **Languages** Python, SQL, C++, R, MATLAB
- **Tools** Amazon Web Services, Google Cloud Platform, Power BI, Anaconda, Docker, GitHub
- **Libraries** NumPy, Matplotlib, OpenCV, PyTorch, TensorFlow, scikit-learn, Keras, pandas, seaborn, cuDF, syft

Experience

- **CarmaCam** — Los Angeles, USA
  Machine Learning Intern — *Aug 2023-Present*
  - Devising two approaches to identify and classify road signs for autonomous vehicles: (1) AutoML on Google Cloud platform, and (2) transfer learning with various architectures (ResNet50, Xception, and InceptionResNetV2).

- **USC Information Technology Services - Office of CISO** — Los Angeles, USA
  Data Scientist — *Feb 2022-May 2023*
  - Led a team of student workers to redesign the framework, predicting risk ratings with an improved **F1-score of 0.91** for 28,000 vendors of USC.
  - Implemented XGBoost model, accomplished a **15% reduction** of false positives in comparison to previous models.
  - Automated the process of alerting vendors of their risk ratings on Power BI, providing data analysis findings to stakeholders with recommendations for mitigating vendor risks. **Cut down 20+ hours** of weekly manual work.

- **Vellore Institute of Technology** — Vellore, India
  Data Science Research Intern — *Nov 2020-Jul 2021*
  - Developed Bi-LSTM and TCN models to anticipate sharp price trends and forecast opening price of a stock, **increased profit revenue by 6%**.
  - Incorporated feature engineering on time series datasets, resulting in **8% improvement in forecast accuracy**.
  - Fine-tuned VADER for sentiment analysis to extract sentiment from reddit articles to predict stock prices. The TCN model with VADER produced a **R squared value of 0.94**.

- **Arista Networks - Reliance Jio** — Mumbai, India
  Machine Learning Intern - Wireless Indoor Localization — *May 2019-Jun 2019*
  - Received theoretical as well as hands-on training on concepts of fingerprinting along with ML algorithms in 1 week.
  - Leveraged k-Nearest Neighbor and Random Forest models to estimate user position in an indoor environment. Using Wi-Fi and inertial sensors yielded positioning as **precise as 2-3 m**.
  - Designed algorithm to apply concepts of RSSI to extract real-time location of client devices operating on access points of WiFi routers placed across work facility with an **accuracy of 0.98.**

Projects

- **Efficiently Detecting COVID-19 in Patient X-ray Images** | *Python, PyTorch, NumPy, Matplotlib*
  - Implemented a multi-stage deep learning system capable of diagnosing a patient with COVID-19 and/or pneumonia given only an X-ray image as input. Employed Unet models, **improved training speeds by a factor of 2**.
  - Deployed this network to the web to get quick results. Achieved **99.3% accuracy and 99.31% F1-score** in the Micronet M3 model.

- **Personal Chat Bot - Study Buddy** | *LLMs, LangChain, OpenAI, Transformers, NLP*
  - Crafted a bespoke chatbot tailored to facilitate interview preparation in the fields of machine learning and data science by ingesting custom personal data.
  - Implemented seamless integration with OpenAI's ChatGPT models, optimizing query interactions through the utilization of a personalized OpenAI API key.

- **Clothing Recommendation System** | *Python, TensorFlow, Keras, NumPy, Matplotlib*
  - Utilized deep learning models, trained on the Deep Fashion dataset, to identify clothing item from images with an **accuracy of 70%** suggested suitable pairing using collaborative filtering.
  - Used collaborative filtering techniques to suggest suitable pairings, providing users with personalized recommendations.

- **Trojan Map** | *C++, OpenCV, Data Structures & Algorithms*
  - Crafted a UI to perform various functions (like Google Maps) on a map of USC and its surroundings. Administered efficient shortest path algorithms Dijkstra and Bellman-Ford. 2-Opt was applied in the Travelling Salesman Problem.
  - Analyzed performance, latency, and runtimes of various algorithms, achieving an **average response time of 5 ms**.