Subharthi Saha

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

Email: subharthi.saha@alumni.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*
- **Vellore Institute of Technology** — Vellore, India
  Bachelor of Technology - Electronics and Communication Engineering **GPA: 8.95/10** — *Jul 2017-Jul 2021*

Relevant Coursework

- Machine Learning
- Deep Learning
- Probability Theory
- Linear Algebra
- Data Structures & Algorithms
- Computer Vision
- Cloud Computing
- Databases
- Statistics
- Natural Language Processing

Technical Skills

- **Languages** Python, SQL, C++, R, MATLAB
- **Tools** AWS, GCP, Power BI, Qlik Anaconda, Docker, GitLab, JIRA, SSMS
- **Libraries** NumPy, Matplotlib, OpenCV, PyTorch, TensorFlow, scikit-learn, Keras, pandas, seaborn, cuDF, syft

Experience

- **Prime Healthcare** — Los Angeles, USA
  Data Scientist — *Feb 2024-Present*
  - Leveraged Prophet (Meta Open Source) to analyze supply utilization and purchasing trends, optimizing reorder points and order quantities. Reduced stockouts by 30%, **saving $1.2 million annually** in expired inventory and capital.
  - Developed a **RAG model with Llama 3.1** to web scrape and categorize healthcare items, suggesting substitutes to enhance purchasing decisions. Improved item substitution **accuracy by 25%**, streamlining procurement and reducing manual categorization efforts.
- **CarmaCam** — Los Angeles, USA
  Software Engineer - Machine Learning Intern — *Aug 2023-Feb 2024*
  - Devised 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*
  - Redesigned the risk prediction framework, achieving improved **F1-score of 0.91** for 28,000 vendors of USC.
  - Implemented **XGBoost** model, accomplished **15% reduction** of false positives, through rigorous **A/B testing**.
  - Automated processes for alerting vendors of their risk ratings on Power BI, provided data analysis findings to stakeholders with recommendations to mitigate vendor risks. **Cut down 20+ hours** of weekly manual work.
- **Vellore Institute of Technology** — Vellore, India
  Data Science Research Intern — *Nov 2020-Jul 2021*
  - Engineered a novel deep-learning model using U-Net to diagnose COVID-19 and pneumonia from X-rays, **improved training speeds by a factor of 2**, reducing diagnosis time, and achieving **low FLOPs** comparable to state-of-the-art models.
  - Deployed this network achieving **99.3% accuracy and 99.31% F1-score** in Micronet M3 model.
- **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

- **Lyft Driver Churn Analysis** | *Python, PySpark, SQL, sklearn, NumPy, Matplotlib, seaborn*
  - Identified churn patterns, setup - guardrail and north star metrics to identify inactive drivers over **7 days**.
  - Estimated **18.48% churn rate**, came up with driver retention strategies by segmenting based on activity patterns and churn indicators.
- **American Sign Language Detection** | *PyTorch, NumPy, Matplotlib, Computer Vision*
  - Utilized ResNet50V2 architecture to predict real-time analysis of hand signs for the disabled. Used Canny Edge Detection technique to pre-process the images and then trained the model on the transformed dataset.
  - Trained model on 87,000 images and yielded **F1 score of 0.99** on test set and real-time analysis.
- **Spotify Song Recommendation Engine** | *Python, sklearn, TensorFlow, Keras, NLP*
  - Merged collaborative, content, and popularity-based filtering techniques for dynamic song suggestions, using weighted averages.
  - Captured semantic meaning of words in lyrics of songs using **word2vec** collaborative filtering techniques to suggest suitable songs, providing users with personalized recommendations with **MAP of 0.83**.