

# Subharthi Saha

LinkedIn: [linkedin.com/in/subharthi-saha/](https://www.linkedin.com/in/subharthi-saha/)

Portfolio: [shubhsaha.com](https://shubhsaha.com)

Email: [subharth@usc.edu](mailto: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  
• **Software Engineer - 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
  - 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
  - Implemented novel efficient deep-learning model to diagnose patients with COVID-19 or pneumonia from X-ray images.
  - Employed Unet encoder-decoder models, **improved training speeds by a factor of 2**, 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, 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*
  - Implemented 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**.
- **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**.