(669) 204-6029
Sunnyvale, California
hamsini.sankaran@berkeley.edu

Hamsini Sankaran

datascienceportfolio/hamsinisankaran
github.com/hamsini1692
linkedin.com/in/hamsini-sankaran

Data/ML Scientist with expertise in data engineering, machine learning, Natural Language Processing (NLP), and data visualization, totaling 6 years of work experience.

Skills

Languages	Python, R, Scala, Golang
Web Dev/Frameworks	TensorFlow,PyTorch,Keras, Django, Flask, React JS, HTML, CSS
Data Analysis	Pandas, NumPy, librosa, Audio Feature analysis, Scipy, Excel
Big Data	Kafka, Akka stream, Elastic Search, Spark, Hadoop, MapReduce, ETL
Database	Postgres, Redis, Mongo, Neo4j, SQL
Visualization	Matplotlib, Seaborn, Plotly, D3.js, Altair,Tableau, Grafana, Splunk, Amcharts, ggplot(R)
Cloud	AWS, GCP, Databricks
Statistics/ML	Scikit-Learn, Hypothesis Testing, A/B Testing, Linear Regression, Logistic Regression, Decision Trees, Random Forest, Xgboost, KNN, SVM, DNN, CNN, GRU RNN, LSTM, Optuna,Transfer Learning, NLP, LLM, Bert, Transfomer, PCA, SSD mobilenet, ensemble, MLP, GBT, Naive Bayes, cross-validation, time-series analysis, text classification, BERT, Transformer
OS/Scripting	Linux, Bash, Windows
Containers/Orchestration	Docker, Kubernetes
Version Control	Git

Education

Master of Information and Data Science, GPA - 4/4 ***University of California, Berkeley*****,** **Expected Graduation: Fall 2024)**
Relevant Courses: Natural Language Processing (NLP), Machine Learning, Data Engineering, Statistics, Data Visualization

Master of Science, Computer Engineering, GPA - 3.42/4 ***San Jose State University*** **Graduation: Fall 2018**

Bachelor of Engineering, Electrical and Electronics, GPA - 3.90/4 ***Anna University*** **Graduation: Fall 2014**

Research Experience

ML Researcher C.H.E.N Labs, UC Berkeley **Jan 2024 – Ongoing**

- Researching the impact of climate on health to explore the complex interplay between environmental factors and public health outcomes.
- Developing predictive models using statistical and machine learning techniques to correlate climate anomalies with health risks.
- **Technologies Used: Python, Tensorflow, PyTorch, GNN**

Data Science Projects

Metastatic Cancer Diagnosis Prediction Women In Data Science Datathon (WIDS), 2024 **Jan 2024 – Feb 2024**

- Spearheaded a three-member team to promote healthcare equity, by pioneering the analysis of oncology data to predict metastatic cancer.
- Achieved a leading AUC score of 80%, showcasing our model's excellence in early diagnosis.
- **Technologies Used: Adasyn, SMOTE, Logistic Regression, Catboost, Random Forest, Xgboost, LightGBM, SVM, Optuna, cross-validation, ensemble, Neural Network, GBT, Xgboost**

Real-Time Prediction of Flight Departure Delays UC Berkeley, CA **Oct 2023 – Dec 2023**

- Led the development of a predictive analytics pipeline, achieving a 52.26% F2 score in forecasting flight delays using historical data.
- Preprocessed a 2+ million-record dataset, performed data cleaning and generated time-frequency features to improve the F2-score.
- Executed hyperparameter tuning via grid search, refining model parameters to bolster precision and recall in predicting flight delays.
- Implemented robust blocked time-series split cross-validation to prevent data leakage, and fine-tuned models using grid search.
- Led data integration efforts to enrich the predictive model with real-time data, enabling more effective trend analysis and decision-making.
- Advocated predictive modeling for cost-efficiency and strategic improvements in airline operations.
- **Technologies Used: Spark, Logistic Regression, Decision Trees, Random Forest, Naive Bayes, MLP, RNN, Ensemble models, Databricks**

ScoreScape - Educational Data Visualization Platform UC Berkeley, CA **Oct 2023 – Dec 2023**

- Developed ScoreScape, a web platform for visualizing educational data including budgets, enrollments, and student performance.
- Utilized U.S. Census Bureau and NAEP data to provide an in-depth view of student achievement in U.S. elementary and secondary schools.
- Enabled educators and policymakers to make data-driven decisions by providing clear, actionable insights.
- Conducted user experience studies with six participants and refined the web interface for an improved user experience based on feedback.
- **Technologies Used: Python, Altair, Tableau, HTML, CSS, Javascript, Bootstrap, Plotly**

Bird Song Classification Using Neural Networks and Machine Learning UC Berkeley, CA **July 2023 – August 2023**

- Spearheaded the use of BirdCLEF 2023 Kaggle data for bird species classification and biodiversity monitoring.
- Used audio augmentation with machine learning models to improve generalization and extract key features like MFCC and chroma.
- Achieved 95% training and 87% testing accuracy using GRU RNN with evaluation based on F1-score metrics.
- **Technologies Used: Python, Tensorflow, Logistic Regression, Random Forest, SVM, Xgboost, CNN, LSTM, GRU RNN**

Revolutionizing Acmet Gourmet Meal (AGM) Delivery with NoSQL Data Magic UC Berkeley, CA **July 2023 – August 2023**

- Led a transformative project at AGM, integrating NoSQL databases for innovative meal delivery solutions.
- Utilized graph algorithms (page rank, community detection and closeness centrality) for optimizing BART delivery networks.
- Implemented customer recommendation systems with personalized route suggestions.
- **Technologies Used: Python, SQL, Neo4j, Graph algorithms, MongoDB, Redis**

Big Budgets? Big Returns? - An Analysis of Film Industry UC Berkeley, CA **March 2023 – April 2023**

- Analyzed The Movie Database (TMDB) movie data for budget-revenue correlations and built regression models on a 30% subsample.
- Demonstrated proficiency in validating assumptions of large samples and classic linear models.
- Led the team and developed regression models to analyze movie revenue with different covariates like movie run time and vote count.
- Identified the best model based on adjusted R^2 and practical significance, showing potential for a 77.2% increase of revenue.
- **Technologies Used: Python, R, Statistics, Linear Regression, OLS**

Decoding Patients' Dilemmas About AI in Healthcare UC Berkeley, CA **March 2023 – April 2023**

- Developed a mixed-methods study to explore patient perspectives on healthcare AI, using surveys and interviews.
- Identified key patient concerns about AI in healthcare, informing patient-centered AI development.
- Designed research methodologies for analyzing patient views on AI, focusing on ethical considerations.
- Proposed strategies to enhance patient trust in AI healthcare technologies.
- **Technologies Used: Research Design, Sampling, AI Ethics**

Work Experience

Software Engineer Walmart eCommerce, Sunnyvale, CA **January 2019–January 2023**

- Designed data-driven solutions using big data and streaming technologies to enhance the Walmart store ecosystem.
- Developed and executed complex SQL queries to uncover core infrastructure insights and correlate data.
- Collaborated with the infrastructure team to predict store networking equipment anomalies and built a scalable portal for 20K users.
- Designed microservices for store infrastructure incident resolution and utilized AIOps to detect and diagnose issues, improving MTTR.

Devops Fellow Engineer Aeris Communications,San Jose, CA **October 2018–December 2018**

- Designed and implemented tools for automating the deployment of IOT applications.
- Created automation to handle disk and memory log errors from Nagios and ELK stack.
- Analyzed server log messages and developed dashboard on Kibana.

Software Engineering Intern Konviv Inc, Berkeley, CA **May 2018–August 2018**

- Developed an AI-based financial management chatbot through K-Means clustering for customer transaction categorization.
- Worked on Bayesian network based recommendation engine to guide the customers for better financial decisions.

Graduate Researcher Tower Foundation of SJSU **May 2017–August 2018**

- Contributed to UAV graffiti removal system development in San Jose, achieving 90% graffiti detection accuracy with SSD Mobilenet.
- Employed inception model to differentiate graffiti and non-graffiti images for drone classification.
- **Technologies used: CNN, SSD mobilenet model, TensorFlow, Python.**

Software Engineer Larsen & Infotech, India **July 2014–June 2016**

- Developed COBOL modules and SQL queries for online travel insurance and grew customer base by 80,000.
- Rectified the production defects and played a major role in improving the latency of the screen designs.

Leadership and Community Engagement

Ambassador at Women In Data Science (WIDS) **Jan 2024–present**

Awards

Employee Of The Month Award: Awarded the best employee of the month for submitting a paper on "Proactive Incident Management leveraging AIOPS techniques" at the internal Walmart conference. **June 2022**