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

Hamsini Sankaran

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

Experienced data science/Machine Learning Researcher at Berkeley AI Research Labs with 6+ years in software engineering (data analytics + data engineering), adept at turning complex data into actionable insights & business gains, proven through significant contributions at Walmart.

Education

Master of Science, Information and Data Science, GPA - 4.0/4.0, ***University of California, Berkeley*****,** **Expected Graduation:** April 2024
Relevant Courses: Natural Language Processing (NLP), Machine Learning, Data Engineering, Statistics, Probability, Data Visualization
Master of Science, Computer Engineering, GPA - 3.42/4, ***San Jose State University*** **Graduated:** Dec 2018
Bachelor of Engineering, Electrical and Electronics, GPA - 3.90/4, ***Anna University*** **Graduated:** June 2014

Skills

Languages	Python, R (R Studio), Scala, Golang
Database	SQL, PostgresSQL, MySQL, Redis, Mongo, Neo4j
Data Science & Analytics	Time-series Analysis, Hypothesis Testing, Calculus, Statistics, Probability, A/B Testing, PCA, Linear Regression, Logistic Regression, Decision Tree, Random Forest, Clustering, Naive Bayes, Predictive modeling, XGBoost, GBM, KNN, Support Vector Machine(SVM), Neural Networks(DNN, CNN, GRU, RNN, LSTM), NLP, ENsemble, Large Language Models(LLM), Transformers, MLP, GBT, Cross-Validation, Text Classification, GPT-4, BERT, Roberta, Llama-2
Frameworks/Web Dev	LangChain, TensorFlow, PyTorch, Keras, Django, Flask, React JS, HTML, CSS
Data Analysis	Pandas, NumPy, Scikit-Learn,Scipy, librosa, Audio Feature analysis, Excel, Jupyter Notebook
Big Data	Apache Spark, Kafka, Akka stream, Elastic Search, Hadoop, MapReduce, ETL
Data Visualization	Matplotlib, Seaborn, Plotly, D3.js, Altair,Tableau, Power BI, Grafana, Splunk, Amcharts, ggplot(R), PowerPoint
Cloud	AWS, Google Cloud Platform (GCP), Databricks
OS/Scripting	Linux, Bash, Windows
Containers/Orchestration	Docker, Kubernetes, CI/CD
Version Control	Git

Research Experience

Machine Learning Researcher, Berkeley AI Research (BAIR) Labs (**Advisor**: Irene Chen) **Dec 2023 – Ongoing**
- Leading a novel research on climate's health impact, exploring environmental-public health links for actionable insights.
- Developing predictive models using statistical and machine learning techniques to correlate climate anomalies with health risks.
- Using high-dimensional, real-world medical data/records from UCSF and HCUP.
- **Technologies Used**: Python, Tensorflow, PyTorch, Graph Neural Networks (GNN)

Data Science Projects

NLP-Driven Analysis of Scientific Articles UC Berkeley, CA **Feb 2024 – Present**
- Developing novel, robust NLP models to classify scientific abstracts using the Elsevier corpus for boosting interdisciplinary research
- Enhanced literature reviews with fine-tuned models for multi-label classification and title generation, achieving a high F1 score of 0.77.
- **Technologies Used**: DAN, CNN, RNN, LSTM, BERT, Roberta, Llama-2/Mistral AI

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

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.
- Processed over 2 million high-dimensional records, cleaned data, and created time-frequency features to boost 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, enhanced predictive models with real-time insights, improved decision-making and efficiency in the airline industry.
- **Technologies Used**: Spark, Spark RDDs, Spark MLib, Logistic Regression, Decision Trees, Random Forest, Naive Bayes, MLP, RNN, Ensemble models, GridSearchCV, 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 – Aug 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 – Aug 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 **Mar 2023 – Apr 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 **Mar 2023 – Apr 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, Analytics - Walmart eCommerce, Sunnyvale, CA **Jan 2019 - Jan 2023**
- Solved large-scale data engineering problems using big data and streaming technologies to enhance the Walmart store ecosystem.
- Designed microservices for store infrastructure incident resolution, utilized AIOps to detect and diagnose issues, **improving MTTR by 2.5x**.
- **E**nhanced efficiency by 42% with an ML predictive system for anomaly detection, scaling support for 20,000 users.
- Led product development through high-quality data sourcing, data-driven design decisions and feedbacks from multiple stakeholders.
- Developed and executed complex SQL queries to uncover core infrastructure insights and correlate data.

Devops Fellow Engineer - Aeris Communications,San Jose, CA **Oct 2018 - Dec 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.

Machine Learning Engineering Intern - Konviv Inc, Berkeley, CA **May 2018 - Aug 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 Deep Learning Researcher - Tower Foundation of SJSU **May 2017 - Aug 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 & Tubro (L&T) Infotech, India **July 2014 - June 2016**
- Developed COBOL modules and SQL queries, driving an 80,000 increase in the online travel insurance customer base.
- 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**