Vivek Reddy

562-879-1191 | vreddy704@berkeley.edu | Los Angeles, CA | Linkedin | GitHub | Portfolio

SUMMARY

Data Scientist with experience applying statistical modeling, experimentation, and machine learning across insurance, healthcare, and consumer analytics. Skilled in Python, SQL, and cloud platforms with a proven record of developing predictive models, feature engineering pipelines, and retrieval/LLM systems. Experienced in risk modeling, fraud detection workflows, and designing generative AI workflows.

EDUCATION

University of California, Berkeley   August 2024
Master of Information and Data Science
Relevant Courses: Research Design and Applications for Data and Analysis, Statistics for Data Science, Fundamentals of Data Engineering, Applied Machine Learning, Natural Language Processing, Machine Learning at Scale, Statistical Methods for Discrete, Time Series, and Panel Data

University of California, Los Angeles (UCLA)   December 2016
B.S., Computational and Systems Biology

PROFESSIONAL EXPERIENCE

Data Engineer (Infrastructure and Analytics) | *Venture Connect*   July 2025 - Present
- Co-developed GitHub Actions workflows for continuous integration and delivery of data pipelines, enabling reliable deployment of new features.
- Contributed to the development of AWS Lambda functions for making updates to data tables on AWS RDS.
- Authored and executed unit tests to validate database update logic, strengthening overall system robustness.

Data Engineer | *Enterprise LLC*   August 2024 – February 2025
- Automated ingestion and orchestration of API data into ETL pipelines, enabling scheduled updates.
- Enhanced a social data scoring algorithm by integrating California AirNow API data, delivering weekly refreshed environmental metrics.

Data Scientist 2 | *MOTER Technologies*   April 2023 – November 2023
- Developed pipeline for extracting road type information from vehicular GPS points to enrich model training data utilizing Docker, Open Street Map, and PostgreSQL.
- Contributed to development of a pipeline leveraging Featuretools library to generate and test many features against important statistical and insurance metrics during iterative model retraining.
- Developed automated SQL-based ETL pipelines with data validation and monitoring dashboards in Amazon Quicksight and Power BI.

Data Scientist 1 | *MOTER Technologies*   May 2022 – April 2023
- Developed predictive models using R, SQL, and Python that used onboard vehicle telemetrics data to develop insurance products for vehicle fleets
- Worked on Python pipelines for automated data ingestion, preprocessing, model fitting, and outputting insurance KPIs.
- Contributed to development of a model and a scoring method that was successfully filed and approved for use in Indiana.
- Actively collaborated in a customer-focused, cross-functional environment to drive projects to fast completion.

Associate Scientist II, Computational Biology | *Neogenomics Laboratories*   December 2020 - April 2022
- Utilized python to implement unsupervised deep learning algorithms to identify unique cellular morphologies.
- Presented results and published conference paper at 2021 American Association of Cancer Research conference https://cancerres.aacrjournals.org/content/81/13_Supplement/154

Associate Scientist I, Computational Biology | *Neogenomics Laboratories*   October 2018 - December 2020
- Utilized Keras-Tensorflow to perform deep learning applications including cell classification, object detection, and tissue segmentation from cancerous tissue images.
- Generated and analyzed statistical reports for global pharmaceutical firms and business clients.
- Created and presented data visualizations using Python libraries such as seaborn and matplotlib to deliver actionable insights to stakeholders.

SELECTED PROJECTS

Assistive Vision Mobile App   August 2025
- Trained a custom YOLOv8 object detection model on 1000+ self-collected images and 5,000+ public dataset samples, achieving real-time, high-precision detection across 20+ grocery item classes.
- Built an AI-powered iOS app that performs on-device detection and selectively routes high-confidence frames to a serverless AWS pipeline (S3 → Lambda → Bedrock) for LLM-based captioning and live voice narration.
- Engineered a low-latency edge-cloud architecture combining CoreML, Vision, and AWS Bedrock, reducing cloud calls by 80% and enabling scalable assistive vision for visually impaired users with <3s end-to-end response time.

Airbnb Housing Recommendation Chatbot (RAG)   August 2024
- Built a retrieval-augmented generation agent with Llama3.1 on AWS Bedrock, FAISS vector store on LangChain for efficient NLP-based retrieval and personalized recommendations.
- Optimized ranking and filtering for high-relevance recommendations.

Natural Language to SQL Query Translation   April 2024
- Finetuned a T5 sequence-to-sequence model to translate natural language questions to SQL given a database schema.
- Achieved an exact string match of 80% and a ROUGE score of 90 using a large cross-domain dataset of over 100 domains utilizing data augmentation, transfer learning, and curriculum learning techniques.

SKILLS & TOOLS

- **Languages**: Python, R, SQL, C#
- **Machine Learning**: Clustering, Logistic Regression, SVM, Random Forest, Neural Networks, PCA, Generative AI (LLMs, RAG , LLM fine-tuning), computer vision (CNNs, LSTMs, YOLOv11), Neo4j, MongoDB

- **Analytics & Modeling**: Experimentation design, Customer Segmentation, Personalization, Metrics Design, EDA, Hypothesis Testing, A/B Testing, Regression, Feature Engineering

- **Visualization**: Tableau, Quicksight, Power BI, Seaborn, ggplot2

- **Cloud & Infrastructure**: AWS (CDK, S3, Athena, Lambda, SageMaker, Bedrock, ECS, Glue), Docker, FastAPI, CI/CD, PySpark, version control

- **ML Frameworks**: Scikit-learn, XGBoost, PyTorch, Keras, LangChain

- **Data Ops**: Docker, Airflow, Spark, ETL design, MLOps, Featuretools