Vijay Ravuri - Data Scientist

San Jose, CA 95118 | (408) 406-6466 | Email | LinkedIn | GitHub | Portfolio

Summary

I am a data scientist with a keen eye for detail and a strong foundation in statistics, coupled with hands-on experience in advanced machine learning techniques. I thrive on exploring data, learning about new domains, and tailoring my technical expertise to meet the unique challenges of each field. With proficiency in Python, R, SQL, and Statistical Modeling, I specialize in developing powerful ML models and providing actionable insights.

Skills

Languages: Python, R, SQL, Git, Scala, Bash, SAS, MongoDB
Statistics: Experimental Design, Time-Series Analysis, Bayesian (MCMC), A/B Testing, Root Cause Analysis, Generalized Linear Models, Multi-Level and Mixed Modeling
Data Science: Jupyter, Pandas, Numpy, Matplotlib, NLP, Tensorflow, Pytorch, LLMs, AWS, Spark, Databricks
Interpersonal Skills: Consulting, Communication, Collaboration, Presentations (technical & non-technical)

Projects

- **Song Recommender:** Classification and Clustering of Song Audio to create a Recommender System
  - Processed and analyzed audio from 8000 songs using the librosa library along with a Convolutional Neural Network and dimension reduction techniques to classify songs by genre.
  - Created a web-app using Streamlit to allow users to find acoustically similar songs to a song of their choice.
- **Election Poll Data Analysis:** Analyzing Poll Accuracy Through Supervised and Unsupervised Methods
  - Used various clustering (K-Means, HDBSCAN) and classification techniques to understand trends between incorrect and correct polls from a set of over 10,000 election polls spanning from 1998 to 2020.
- **NLP Text Classification:** Distinguishing Between Real and AI-Generated Reddit Comments
  - Applied PRAW and OpenAI APIs to gather real and AI-generated reddit responses to real posts.
  - Trained, tuned, and compared the performance of various classification methods (Ensemble Tree Models, Boosting, SVM) and identified the optimal model across relevant metrics (F-1 Score, Precision, Recall)

Experience

CUHSD | Varsity Badminton Coach / Python Workshop Lead/Mentor | San Jose, CA | 2020-2024
- Managed a team of 48 high school students, including scheduling, creating practice plans, and maintaining clear communication lines among players, parents, and school administration.
- Mentored players with a focus on individual long-term goals, while fostering teamwork and collaboration to achieve team-wide goals
- Conducted a 90-minute workshop to teach students with low coding and technical experience CRUD through MongoDB.

Survey Consultant | Behavioral Health Department | San Luis Obispo, CA. | 10/2017-12/2017
- Utilized R and SAS to run hypothesis tests and create data visualizations to address client's survey questions and concerns about impact about department's Diversity, Equity, and Inclusion training.
- Wrote a quantitative analysis of the survey, including research methodology, results, and key takeaways for department officials. Presented findings and answered questions from department officials regarding our results and analysis

Education

General Assembly | Data Science Immersive | Remote | 11/2023

California Polytechnic State University | Bachelor of Science in Statistics | San Luis Obispo, CA
- **Related coursework**: Statistical Programing (R), Probability Theory, Estimation & Sampling Theory, Survival Analysis, Categorical Data Analysis, Object-Oriented Programming & Design, Data Structures & Algorithms