Vijay Ravuri - Data Scientist

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

Summary

I am a detail-oriented statistician and data scientist driven by a passion for deciphering the intricate relationship between trends in data and creating actionable change. I view statistics as a powerful and versatile tool that is uniquely suited for application across and between diverse domains. My goal is to combine creative problem solving, adaptability, along with a solid foundation in theoretical and applied statistics to experience a wide range of disciplines while gaining deeper insights and perspectives with the aim of making a meaningful impact in the world.

Skills

Languages: Python, R, SQL, Git, Scala, Bash, SAS
Statistics: Multi-level & Mixed Modeling, Experimental Design, Time-Series Analysis, Bayesian (MCMC)
Data Science: Jupyter, Pandas, Numpy, Seaborn, Matplotlib, ggplot, NLP, Tensorflow, Pytorch, caret, mlr
Interpersonal Skills: Consulting, Communication, Collaboration, Presentations (technical & non-technical)

Projects

- **Song Analysis:** Classification and Clustering of Song Audio to create a Recommender System
    - o Processed and analyzed audio from 8000 songs using the librosa library along with Neural Networks and dimension reduction techniques to classify songs by genre.
    - o Created a web-app using Streamlit to allow users to find acoustically similar songs to a song of their choice.
- **Poll Data Analysis:** Analyzing Poll Accuracy Through Supervised and Unsupervised Methods
    - o 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
    - o Applied PRAW and OpenAI APIs to gather real and AI-generated reddit responses to real posts.
    - o Trained, tuned, and compared the performance of various classification methods (Ensemble Tree Models, Boosting, SVM) and identified the optimal model.

Experience

GENERAL ASSEMBLY | DATA SCIENCE FELLOW | REMOTE | 11/2023
- Learned and applied machine learning and data visualization techniques using industry standard platforms.
    - o Jupyter, Scala, SQL, Visual Studio Code, Google Colab, Databricks
- Completed projects with an emphasis on real-world impact and performing analysis for tangible results.
- Relayed results of projects and key findings from analyses with an emphasis on non-technical audiences

SURVEY CONSULTANT | BEHAVIORAL HEALTH DEPARTMENT | SAN LUIS OBISPO, CA. | 10/2018-12/2018
- Administered, analyzed, and reported on a survey to assess the cultural competency of the department.
- 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