SHANNON RUMSEY

San Jose, CA • shannonalmyrumsey@gmail.com • https://www.linkedin.com/in/shannon-rumsey/

EDUCATION

University of California, Santa Barbara — Santa Barbara, CA
B.S. in Statistics and Data Science — September 2021-December 2023
Data Science Club
Recipient of National Science Foundation's Harnessing the Data Revolution Data Science Corps Award

Santa Barbara City College — Santa Barbara, CA
A.A. in Liberal Arts-Science: Science & Math — August 2019-May 2021
President's Honor Roll

EXPERIENCE

Postbaccalaureate Research Assistant, CPLS Lab, Remote — June 2023-Present
- Validate and build upon 15 prior linguistics research papers of the Functional Load Hypothesis to determine potential predictors through a mixed effects logistic regression model
- Quantify the extent to which minimal pairs may be confused with each other in context using Word2Vec embeddings trained on hand-picked, preprocessed corpora
- Examine the effects of different window sizes and part of speech tags to explore whether syntactic or semantic context are more important in determining potential confusability

Capstone Researcher, Evidation Health, Santa Barbara, CA — January 2023-June 2023
- Directed a team of 5 to analyze remotely collected wearables data, such as an individual's resting heart rate and caloric expenditure, to predict the outcome of participant Respiratory Viral Infection lab tests
- Explored different techniques of data imputation (MICE) and dimension reduction, such as UMAP (Uniform Manifold Approximation and Projection) and t-SNE (t-distributed stochastic neighbor embedding)
- Presented at the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

Undergraduate Data Science Fellow, Central Coast Data Science Fellowship, Santa Barbara, CA — September 2022-June 2023
- Led outreach programs aimed at inspiring students and promoting departmental courses to a diverse student body
- Organized and presented at the Department of Probability and Statistics Project Showcase

RESEARCH PROJECTS

Independent Researcher, Medical Transcription Simplification Using BERT, Santa Barbara, CA — April 2023-June 2023
- Discovered an increase in readability of medical transcriptions by reducing Flesch-Kincaid metric, making information contained more accessible to those without domain expertise
- Computed cosine similarities between BioBert embeddings derived from Harvard Health medical definitions and context dependent terminology found in the transcriptions to discover synonym substitutions
- Employed Bert2Bert to summarize and further refine the overall documents

Independent Researcher, Yelp Classification Modeling, Santa Barbara, CA — September 2022-December 2022
- Assessed the relationship between geographic attributes and review sentiment to aid in business development strategies
- Trained 4 machine learning methods to identify key predictors for defining a business's location within the Santa Barbara county

SKILLS

Disciplines	Computational Linguistics, Bayesian Data Analysis, Statistics and Probability, Regression Analysis, Machine Learning, Natural Language Processing, Statistical Data Science
Languages	Python, R, SQL, SAS, HTML, Quarto