

Komal Sairam Reddy Bhimireddy

(669)-204-6802 | komalbhimireddy@gmail.com | Portfolio | LinkedIn | GitHub | San Francisco, USA

SUMMARY

Machine Learning Scientist with expertise in LLMs, NLP, and medical speech applications, experienced in building end-to-end pipelines for transcription, summarization, and predictive modeling. Skilled in fine-tuning and evaluating foundation models and applying statistical modeling and causal inference to extract actionable insights. Proven ability to combine ML experimentation with data science rigor to deliver interpretable, production-ready solutions for healthcare.

EDUCATION

San Francisco State University, MS in Statistical Data Science, GPA: 3.66 **May 2025**
Coursework: Data Mining, Probability & Statistics, Advanced Probability Models, Experimental Design
Computational Statistics, Statistical & Machine Learning, Multivariate Statistical Methods

EXPERIENCE

ML Engineer Intern **July 2025 – September 2025**
Suki AI *Redwood City, CA*

- Developed and optimized an ASR data pipeline for medical speech, covering ingestion, deduplication, augmentation, and Whisper fine-tuning to improve accuracy and reproducibility.
- Implemented targeted augmentation strategies (noise injection, TTS synthesis, pitch/speed variation) to simulate clinical acoustic variability and improve downstream NLP generalization.
- Fine-tuned Whisper-Medium on optimized datasets, achieving an 8.5% WER reduction and 6.7% recall improvement over baseline, enhancing robustness to noisy, real-world clinician speech.
- Designed LLM-based evaluation workflows for ASR transcripts, improving semantic understanding, summarization quality, and intent recognition in clinical voice data.

PROJECTS

BRFSS SMART Analysis: Binge Drinking & Mental Distress | Code **January 2025 - May 2025**

- Constructed a analytic dataset of 726,000+ BRFSS respondents by standardizing variable coding; defining consistent outcomes and predictors. Preserved survey design weights to ensure valid estimation under complex sampling design.
- Implemented a stepwise logistic regression to evaluate the association between binge drinking and frequent mental distress. Sequentially added social, behavioral, health covariates to show how confounding shaped effect estimates.
- Identified socioeconomic status and smoking as key confounders, with fully adjusted model showing binge drinking as a modest independent risk factor (OR \approx 1.04). Improved model fit with a 100k reduction in AIC and R^2 rising to 0.110.

Enhancing E-commerce Experience through Sentiment Analysis | Code **February 2024 – April 2024**

- Developed a sentiment analysis model for 205,052 e-commerce reviews, leveraging TF-IDF vectorization and machine learning models (SVM, Random Forest, Gradient Boosting) for automated sentiment classification.
- Experimented with data balancing techniques (Tomek Links, SMOTE, Random Under-Sampling) and found that SVM naturally handled class imbalance, making additional resampling unnecessary.
- Achieved an F1 Score of 0.9360 (+3%) and Recall of 0.9419 (+4.6%) using SVM, proving sampling was unnecessary and reinforcing the need to test pre-processing techniques before applying them.

PUBLICATIONS

Forecasting Gold Returns Volatility Over 1258–2023: The Role of Moments | Link **September 2025**
Peer-reviewed publication in Applied Stochastic Models in Business and Industry, 2025

- Applied Bayesian time-varying quantile regressions on 766 years of gold return data, demonstrating that tail risks and higher-order moments substantially enhance volatility forecasts over autoregressive benchmarks.

SKILLS

Programming: Python, R, SQL, Shell Scripting, Git
Machine Learning & NLP: LLMs, Transformers, Text Classification, Sentiment Analysis, Topic/ Predictive Modeling
Frameworks: PyTorch, TensorFlow, Hugging Face, Scikit-learn, Pandas, NumPy, OpenAI APIs
Data Science & Analytics: A/B Testing, Statistical Inference, Regression Modeling, Experiment Design, Causal Inference, Model Interpretability, Time Series Forecasting
ML Engineering: Data Pipeline Automation, Feature Engineering, CI/CD, Google Cloud Platform (GCP), Model Benchmarking, Error Analysis

EXTRACURRICULAR ACTIVITIES

Math Tutor | San Francisco State University February 2024 – May 2025
Graduate Teaching Associate | San Francisco State University August 2024 – May 2025