# Komal Sairam Reddy Bhimireddy

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

## SUMMARY

ML Engineer with experience in Python, SQL, and cloud-based development, skilled in building scalable pipelines and systems. Proficient in data preprocessing, automation, and model evaluation with hands-on expertise in ASR data workflows. Focused on leveraging AI and infrastructure practices to deliver reliable and efficient solutions.

## EDUCATION

**San Francisco State University, MS in Statistical Data Science, GPA: 3.66** **May 2025**

**Coursework:** Data Mining, Probability and Statistics, Advanced Probability Models, Computational Statistics, Statistical and Machine Learning, Multivariate Statistical Methods

## EXPERIENCE

**ML Engineer Intern** | **Suki AI** **July 2025 – September 2025**

- Designed and implemented a complete ASR data pipeline from raw data ingestion through EDA, deduplication, targeted augmentation, and Whisper fine-tuning to improve model accuracy and dataset reliability.
- Removed 1,087 hidden audio duplicates via waveform hashing and generated 1,700+ synthetic commands using TTS engines to balance gender representation and underrepresented semantic clusters.
- Fine-tuned Whisper-Medium on the optimized dataset, achieving an 8.5% reduction in WER and a 6.7% recall improvement over baseline, establishing a stronger ASR foundation for clinician workflows.

## PROJECTS

**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.

**Music Recommendation System** | Code **October 2023 – December 2023**

- Engineered a music recommendation system using Spotify API data, applying machine learning models to predict song popularity and enhance recommendation accuracy.
- Developed and fine-tuned regression & classification models (Random Forest Regressor: $R^2 = 0.269$, Random Forest Classifier: Accuracy = 0.671) to estimate song popularity and improve ranking of recommended tracks.
- Implemented a content-based recommendation engine using Cosine Similarity, analyzing 5,333 songs to personalize suggestions, demonstrating practical application of machine learning in real-world music analytics.

## PUBLICATIONS

**Forecasting Gold Returns Volatility Over 1258–2023: The Role of Moments**| Link **September 2025**

*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
**Cloud & Infrastructure:** Google Cloud Platform, Kubernetes, Workflow Automation, Data Pipeline Development
**Machine Learning:** Supervised & Unsupervised Learning, Model Fine-tuning, TTS-based Data Augmentation, Feature Engineering, Hyperparameter Optimization, Model Evaluation
**Statistical Methods:** Bayesian Inference, Quantile Regression, Multivariate Analysis, Time Series Forecasting

## EXTRACURRICULAR ACTIVITIES

Math Tutor | Tutoring & Academic Support Center February 2024 – May 2025
Graduate Teaching Associate | San Francisco State University August 2024 – May 2025