MUKHIL GUNA

Los Angeles, CA | mukhil@g.ucla.edu | (408) 416-7199 | linkedin.com/in/mukhilguna

EDUCATION

University of California, Los Angeles || Los Angeles, CA — Dec 2025

B.S. Statistics & Data Science

Relevant Coursework: Probability and Distributions, Data Analysis and Regression, Monte Carlo Methods / Simulation, Linear Algebra, Bayesian Statistics

SKILLS

- **Python** (pandas, sklearn, TensorFlow, PySpark, PyTorch, NumPy, matplotlib, seaborn), **SQL** (MySQL, SQL Server), **R** (dplyr, tidyr, ggplot2), Tableau, Power BI, Spreadsheets

EXPERIENCE

NBA Performance Prediction Model: Machine Learning Group Project — 09/2024 – 12/2024

- Developed an ML pipeline that achieved 71% prediction accuracy in NBA game outcomes, ranking in the 90th percentile for accuracy among comparable published models
- Orchestrated model evaluation across 5 ML algorithms (Logistic Regression, LDA, Random Forest, Gradient Boosting, and SVM) using k-fold CV and GridSearchCV, optimizing for both accuracy and interpretability
- Implemented time-weighted feature engineering across 20+ statistical metrics, improving model performance by 3%
- Created exponential weighting functions to capture temporal patterns in sports analytics, enhancing team performance prediction

Music Data Analysis Using PySpark: Personal Project — 07/2024 – 08/2024

- Created a data processing pipeline to analyze 60,000+ unique songs on Spotify, implementing SQL queries, PCA, and K-means clustering to reveal specific insights into artist popularity and music taste evolution
- Developed a cosine similarity algorithm for song recommendations, processing the dataset to find similar tracks, while integrating the Spotify API
- Created visualizations using seaborn to effectively communicate complex trends in music popularity and audio characteristics

Sign Language Recognition using CNN: National Data Science Corps — 01/2024 – 03/2024

- Developed CNN models for sign language recognition using TensorFlow and PyTorch, achieving 99.93% validation accuracy and 92.21% test accuracy by utilizing K-fold cross validation and batch processing over 28,000 images
- Optimized data preprocessing and augmentation workflows, with the use of normalization, reshaping, and Image Data Generator to enhance training efficiency, mitigating overfitting and improving model generalization

Predicting Solar Radiation Levels using Multiple Linear Regression: — 01/2024 – 03/2024

Presentation

- Developed a power transformed multiple linear regression model to predict solar radiation levels, achieving an R^2 value of 0.6404, and optimized model performance using AIC, AICC, BIC and VIF metrics
- Performed data preprocessing, including time series averaging, autocorrelation reduction, and multicollinearity analysis, improving the model's predictive accuracy aligning predictions with environmental patterns such as Hawaii's trade winds