

PRASHANTH KUMAR GUNDA

☎ 954-560-2208 ✉ gundapr@mail.uc.edu [in LinkedIn](#) [@ GitHub](#)

Data Science professional with 4 years of work experience in statistical analysis & machine learning using **Python, R, & MLOps/AIOps**. Skilled in deep learning with **PyTorch, Keras & TensorFlow**. Proficient in NLP (LLMs, prompt engineering, RAG chatbots), cloud deployment (**Azure, AWS, GCP**) & data visualization tools **Tableau & Power BI**.

WORK EXPERIENCE

Data Scientist Intern

April 2024- June 2024

GEN1E Life Sciences, INC

Palo Alto, SanFrancisco

- Developed ML pipeline using MLOps to identify ARDS patient endotypes based on readily available clinical parameters.
- Identified the base-line parameters by data manipulation techniques using libraries such as **Pandas and Numpy** and performed advanced EDA on datasets of over **300,000** patients to extract features for endotype classification.
- Utilized A/B testing to compare the performance of various machine learning models. Achieved the highest **ROC-AUC** score of **0.98** with a Support Vector Machine (SVM) model.
- Used **Logistic regression** to check impact of HTE classification and endotype classification on clinical outcomes finding that only endotype classification is significant, thus validating the model's effectiveness.

Graduate Assistant

Aug 2023-Dec 2023

University of Cincinnati

Cincinnati, Ohio

- Designed a novel pipeline using **Clinical BERT** for advanced **Named Entity Recognition (NER)** and **sentiment analysis**, optimizing the extraction of social determinants from diabetes patient records
- Analyzed a vast dataset of **10M** electronic health records (EHRs) using **SQL, PyTorch, & high-performance computing**, revealing SDOH for diabetes.
- Enhanced Clinical BERT with **Parameter Efficient fine-tuning (PEFT)** and **Low-rank adaptation (LoRA)** for NER and sentiment analysis leading to **10%** increase in model accuracy.
- Built a chatbot utilizing **RAG and LangChain** techniques and **streamlit** to fetch patient records enabling user interaction about social determinants. This enhanced targeted care, reduced time to access patient insights by **25%**.

Data Analyst

Mar 2022- July 2023

Infosys

Hyderabad, India

Machine Learning Analytics and Modeling for Detecting Insurance Fraud

- Developed a model pipeline to identify fraudulent insurance claims using a database containing more than **100M** records.
- Conducted exploratory data analysis (EDA), imputed missing values, and applied categorical encoders such as **One-Hot Encoding and Target Encoding** for categorical features using **pandas and scikit-learn** libraries.
- Built fine-tuned multiple models, including **Decision Trees, Random Forest, AdaBoost, Gradient Boosting, XGBoost, CatBoost, and statistical models**, using **Grid Search CV** for hyperparameter optimization.
- Employed ensemble techniques to combine these models, resulting in a final ensemble model that achieved an accuracy of **85%**, with an adjusted R-squared value of **0.78**.

Forecasting Insurance Claim Amounts Using ARIMA Models

- Conducted transformation of non-stationary insurance claim amount data into stationary data by using rolling statistics.
- Visualized data trends and seasonality through **correlation and autocorrelation plots** by **ACF and PACF**, providing actionable insights that reduced forecasting errors by **20%**.
- Built **ARIMA model** for accurate forecasting of claim amounts using **statsmodels**, improving forecast accuracy by **30%**, thus enabling more precise decision making.

Data Analyst

Aug 2021-Jan 2021

BYJU'S

Bengaluru, India

- Developed a web platform using **HTML, CSS, and JavaScript** to capture and log user engagement times with Maths content, such as time spent and interaction patterns.
- Extracted and processed the engagement data using **SQL**, applied **K-Means clustering** to identify 5 distinct patterns (**silhouette score of 0.84**), & optimized content delivery strategies, resulting in a **20%** increase in student engagement.

Research Associate

July 2020- July 2021

Project: Machine Learning Application in Cavitation Induced Vortex Dynamics

IIT Hyderabad, India

- Used SVM algorithm to analyze the vibration data and understand formation of vortices.
- Built supervised machine learning model that achieved **86%** accuracy in predicting optimal cluster membership under varying environmental conditions like liquid density, temperature and viscosity on bubble interactions.

EDUCATION

University of Cincinnati, Carl H. Lindner College of Business

Expected: Aug 2024

Master of Science in Business Analytics

CGPA: 3.7

Indian Institute of Technology, Hyderabad

Aug 2021

Master of Technology in Engineering

GPA: 4.0