Prashanth Kumar Gunda

(954) 560-2208 • gundapr@mail.uc.edu • Cincinnati, Ohio • LinkedIn • GitHub

EDUCATION

University of Cincinnati, Carl H. Lindner College of Business Cincinnati, Ohio
Master of Science in Business Analytics

Indian Institute of Technology, Hyderabad Hyderabad, India
Master of Technology in Engineering

SKILLS

- **Certifications**: Azure Enterprise Data Analyst Associate, Azure Fundamentals, Azure Data Engineer Associate
- **Analytical, and Visualization tools:** SQL, R, NoSQL, Excel ,Tableau and PowerBI
- **Programming tools:** Python, MATLAB, Java, Scala, Streamlit, HTML, CSS, C, C++
- **Deep Learning Libraries, and Architectures**: PyTorch, Keras, TensorFlow, Transformers, CNN, VGG16
- **Generative AI :** NLP, NLTK,spaCy, LLMs (GPT, BERT, T5), RAG, Prompt Engineering, LangChain, RAGAS
- **Machine Learning:** Regression, Clustering, Classification, Gradient Boosting, Neural Networks, Sklearn, Scipy

WORK EXPERIENCE

Data Scientist Intern. Apr 2024- Jun 2024
GEN1E Life Sciences, INC San Francisco, CA

- Analyzed **300K unique** patient records and used readily available clinical parameters to build an **SVM** model classifying **ARDS** endotypes, projecting a **30% reduction** in mortality rate for a major hospital chain.
- Developed **ML pipeline** using **MLOps** to identify ARDS patient endotypes and utilized **GridSearchCV** for hyperparameter tuning, achieving the highest **ROC-AUC** score of **0.74** with an **SVM model**.
- Built a self-servicing **Power BI** dashboard to **automate** gene-drug-disease mapping, allowing for quick visualization of genes, reducing data retrieval time from **minutes to seconds,** compared to **graphite** library in **R**.

Graduate Assistant Aug 2023-Dec 2023
University of Cincinnati Cincinnati, OH

- Analyzed a dataset of **10M** electronic health records (EHRs) using **SQL and PyTorch**, and designed a **novel** pipeline using **Clinical BERT** for advanced **NLP** analysis, optimizing the extraction of social determinants for diabetes.
- Optimized **Clinical BERT** with **PEFT** & **LoRA** leading to **30% decrease** in computational time.
- Built a chatbot utilizing **RAG** and **LangChain** techniques and **Streamlit** to fetch patient records enabling user interaction about social determinants **reducing time** to access patient insights by **25%.**

Data Analyst Mar 2022 – Jul 2023
Infosys Hyderabad, India

- Orchestrated a **pipeline** for identifying **fraudulent** insurance claims using a database exceeding **100M** records, with data curated and compressed using **Hadoop** and **Hive**, thus reducing data processing time by **40%** .
- Utilized **ETL** tools and integrated **SAS** for advanced data processing and preprocessing.
- Conducted **EDA** using Python libraries, and employed **Spark** with **Hadoop** for **distributed data** processing.
- Implemented **ensemble** techniques on **Random Forest** and **XGBoost** to achieve a model of **78%** accuracy in **fraud detection**, contributing to a projected **$1.5M** increase in annual company revenue.
- Deployed the final model using **AWS,** & integrated **MLflow** for monitoring, increasing efficiency **by 35 %.**
- Built an **ARIMA** model for forecasting claim amounts using **statsmodels**, achieving forecast accuracy upto **70%** .
- Employed **Tableau** to forecast data and **KPIs** like loss ratio and pure premium, thus saved the firm **$500,000.**

Data Analyst Jul 2021 –Jan 2022
BYJU'S Bengaluru, India

- Designed a web platform using **HTML**, **CSS**, and **JavaScript** to capture and log user engagement times, tracking over **50K** interaction patterns &, employed **A/B testing** to optimize interface elements.
- Processed the engagement data using **SQL**, applied **K-Means clustering** to identify **5 distinct patterns** (**silhouette score of 0.84**), and optimized content delivery strategies, leading to **20%** increase in engagement.

Research Associate Jul 2020 – Jul 2021
Indian Institute of Technology, Hyderabad Hyderabad, India

- Developed unsupervised model (**K-means**) that achieved **86%** accuracy in predicting optimal cluster membership under varying environmental conditions like liquid density, temperature and viscosity on bubble interactions.