Rithvik Krishna Donnipadu

+1-(650)-433-6877 | rithvikdonnipadu@gmail.com | Linkedin | GitHub | Medium | San Francisco, CA

Education

University of San Francisco | MS in Data Science | San Francisco, CA *2023 - 2024*
Relevant Coursework: Regression, Time Series Analysis, Probability, Statistics, A/B Testing, Data Structures, Algorithms, Relational Databases (SQL), Machine Learning, Distributed Computing, Data Acquisition, Deep Learning, NLP, LLMs, Data Ethics

BITS Pilani | B.E in Electrical Engineering | Hyderabad, India *2017 - 2021*
Relevant Coursework: Calculus, Algebra, Optimization, Operating Systems, Human Computer Interaction, Object Oriented Programming, Computer Architecture

Technical Skills

Programming: SQL, Python, R, Git, Linux, PySpark, Hive, C, C++, Java
Technologies/Frameworks: Tableau, GCP, AWS, Snowflake, dbt, MongoDB, Postgres, Airflow, Streamlit, Spark, Hadoop, Github, UC4, Selenium, Teradata, BigQuery, Excel
Packages: Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, LightGBM, Pandas, NumPy, SciPy, StatsModels, Matplotlib, Seaborn, Plotly, Spacy, NLTK, Gensim
Machine Learning: Regression, Classification, Recommendations (Content based and Collaborative filtering), Anomaly Detection, Deep Learning-NLP, Transformers, Multi Armed Bandits, Reinforcement Learning, Large Language Models (LLM), Ensemble Models

Work Experience

Stanford University | Data Scientist, Part-Time | Palo Alto, CA Oct 2023 - Present
Glaucoma Disease Prediction

- Created a cohort of glaucoma patients by querying through more than **1 million** patient encounters and **terabytes of data** in Big Query using GCP. This project aims at improving doctor workload by saving almost **3,000 hours/year**
- Performed feature engineering to extrapolate **~13,000 patients'** data, which led to **500+ features.** Used FPCA (***Principal Component Analysis***) to construct a longitudinal disease progression model. With an 80:20 train-test split, the prediction of Glaucoma trajectories 10 years into the future using FPCA and 1 year of historical data yielded a model with **R-squared** value of **74%**.
- On track for having a publishable paper by **Aug 2024 as a 1st author** on predicting glaucoma progression using **FPCA/deep learning** models that incorporates **Terabytes** of both EHR and structured data through data fusion architectures.

PayPal | Data Scientist | Bengaluru, India Feb 2021 - June 2023
Consumer KPI enhancement recommendations

- Spearheaded **anomaly detection** based predictive modeling initiatives on PayPal consumer data using **XGBoost**, to optimize the activation rates KPI across diverse customer cohorts by ~ **5%**. Led the design and implementation of scalable **Machine Learning** pipelines in production, utilizing frameworks such as PyTorch, and Scikit-learn.
- Successfully migrated an on-premise **Hadoop** cluster framework to **GCP**, achieving a notable **25% reduction** in processing time and enhancing scalability. Expanded anomaly detection capabilities to multiple business verticals within PayPal, demonstrating adaptability and cross-functional leadership.

Churn analytics and reporting for strategic decision-making

- Maintained a robust churn model for EMEA and APAC markets, supporting a customer base valued at **$1B**.
- Presented detailed monthly reports to the finance team, providing actionable **insights** into churn trends and potential revenue impacts.
- Conducted **A/B testing** on churn prevention initiatives aimed at reducing telephone traffic to PayPal call centers, resulting in a significant **4% reduction** in overhead costs and streamlining customer support operations.

Tableau dashboard for actionable insights

- Developed a comprehensive product features usage **dashboard**, which facilitated a rapid increase in adoption of the Digital Wallet product, scaling from **1% on Day 1 to ~95%** within a span of 3 months.
- Used **tableau** to visualize over **1 million rows** of data, and processed data through more than **100 queries on BigQuery**. Additionally, automated email alerts were set up to disseminate insights to approximately 200 stakeholders.
- Executed the seamless migration of over **50 tables and 100 SQL queries** from **Teradata** platform to **Snowflake**, ensuring uninterrupted production operations.

Select Academic Projects

Recommendation System - Developed an *article recommendation system* using Stanford GloVe word vectors and euclidean similarity which was trained on a dump of wikipedia data. (GitHub, Medium)
ML from scratch - Implemented Decision Trees, Random Forest, Boosting, Adagrad, k-means clustering from scratch in Python