AMARTYA YALLA

858-744-3130 | amartya1268@gmail.com | LinkedIn | Portfolio

PROFILE SUMMARY

- Detail-oriented **Data Analyst with 5 years of experience** in **data engineering, Statistical analytics, and machine learning**, specializing in **healthcare and financial risk analytics**, with strong proficiency in **SAP**, **Python** (Pandas, NumPy, Scikit-learn) **and SQL** (PostgreSQL, MySQL, SQL Server, BigQuery).
- **Strong SQL expertise**, skilled in **query optimization, indexing, partitioning, window functions, and data modeling** to efficiently manage and analyze large-scale datasets.
- **Cloud & ETL proficiency**, designing and managing **scalable data pipelines in Google Cloud** (BigQuery, Cloud SQL)**, AWS** (S3, RDS), **Databricks, Apache Airflow, and SSIS** to automate data workflows.
- **Machine learning expertise**, applying **predictive modeling** (Random Forest, XGBoost, LSTM), **deep learning (**BERT, CNNs)**, and anomaly detection** for fraud detection and risk assessment.
- **Stakeholder collaboration**, working closely with business leaders, compliance officers, engineers, and data teams to develop data-driven solutions that optimize operations and enhance decision-making.

WORK EXPERIENCE

Data Analyst, Molina Healthcare Inc   **11/2023 – present**

- **Engineered ETL pipelines** using **SQL Server Integration Services (SSIS), Python (Airflow, Pandas, NumPy), and Databricks**, reducing data processing time by 40% and ensuring **data integrity**.
- Developed a **predictive risk scoring system** for patient readmission using **XGBoost and Logistic Regression**, integrating model outputs into Power BI dashboards for clinical teams to identify high-risk cases in real-time.
- **Applied statistical hypothesis testing (Chi-Square, ANOVA, and T-Tests)** to validate patient readmission risk models, ensuring a **95% confidence level** in predictive outcomes.
- Implemented **anomaly detection algorithms (Isolation Forest, DBSCAN)** to flag fraudulent claims and unusual billing patterns and **Bayesian inference and Monte Carlo simulations** to quantify uncertainty in claim fraud detection, improving fraud risk assessment accuracy by **20%**
- **Designed deep learning-based medical text classification models (BERT, LSTM, TF-IDF + Naïve Bayes) using TensorFlow & Scikit-learn**, automating **provider notes classification**, increasing coding efficiency by 30%.
- **Implemented geospatial risk analysis using PostGIS, ArcGIS & BigQuery GIS**, mapping **healthcare facility access gaps** and optimizing service coverage, increasing accessibility by 25%.
- Integrated **SAP HANA** with Google Cloud BigQuery for real-time claims data synchronization, enabling advanced analytics and reducing reporting latency by 40% through optimized data federation and **SAP OData service utilization**.
- Streamlined **HIPAA and CMS** compliance reporting by developing automated workflows by **DAX scripting** on BigQuery, reducing manual effort by 60% and accelerating regulatory insight delivery across 5TB+ claims data.
- Collaborated with **cross-functional teams** (cloud engineers, compliance officers, **business stakeholders**) to deploy cloud-based reporting solutions on **GCP**, ensuring seamless data accessibility and governance compliance.

Impact: Reduced pipeline failures by 70%, improved predictive analytics accuracy, and enhanced fraud detection mechanisms for healthcare claims.

Data Analyst, Omega Healthcare   **04/2019 – 05/2022**

- **Built a Time Series Forecasting Model (Prophet, ARIMA, LSTM) in Python**, predicting **hospital admission trends**, increasing resource allocation efficiency by 25%.

- Developed an **AI-driven fraud detection pipeline** using **Random Forest, CatBoost, and Isolation Forest**, reducing fraudulent claim losses by 15% while ensuring model adaptability to new fraud patterns.
- Engineered **geospatial** fraud detection models using **PostGIS (ST_Buffer, ST_Intersects)**, detecting duplicate provider locations, increasing fraud detection accuracy by 15%.
- **Applied A/B testing frameworks (Multi-Armed Bandit, T-tests, Bayesian Optimization)** to optimize reimbursement strategies, increasing revenue cycle efficiency by 18%.
- Automated **Excel-based compliance reporting** using **Python (OpenPyXL, XlsxWriter), VBA macros, and Power Query**, ensuring 100% accuracy in regulatory filings.
- **Optimized MySQL query execution (partitioning, indexing, query caching)** for large-scale claim datasets (1B+ rows), improving transactional performance by 40%.
- Designed **Power BI dashboards with REST API integrations (JSON)** for real-time claim denial analysis, providing finance teams with on-demand insights to track claim approval trends.
- **Partnered with finance, compliance, and fraud teams** to optimize reimbursement models using ML-powered claims forecasting, reducing financial risks.

Impact: Increased claims processing efficiency by 35%, reduced fraudulent claim losses by 15%, and improved predictive analytics-driven revenue strategies.

TECHNICAL SKILLS

Programming & Data Engineering: SQL (PostgreSQL, MySQL, SQL Server, BigQuery), Python (Pandas, NumPy, Scikit-learn, TensorFlow), R (Tidyverse, ggplot2), Java

Cloud & ETL: Google Cloud (BigQuery, Cloud SQL), AWS (S3, RDS), Databricks, Apache Airflow, SSIS, PySpark

Machine Learning & AI: Predictive Modeling (Random Forest, XGBoost, Logistic Regression), Deep Learning (BERT, LSTM, CNNs), NLP (spaCy, TF-IDF), Anomaly Detection (Isolation Forest, DBSCAN)

Business Intelligence & Visualization: Power BI (DAX, M Query), Tableau, Advanced Excel (Pivot Tables, VLOOKUP, Macros), SSIS, SPSS, Lucid Chart

Geospatial & Analytics: ArcGIS, QGIS, Geopandas, PostGIS, Simpson's paradox, Regression Analysis

DevOps & Version Control: GitHub, Docker, Kubernetes, CI/CD, JIRA, Agile,

EDUCATION

Master's: **Data Analytics**	**10/2024**
Alliant International University	San Diego, USA
Bachelor's: **Computer science**	**12/2020**
Vellore Institute of Technology	Vellore, India

PUBLICATION & CERTIFICATIONS

- **Machine Learning-Based Anomaly Intrusion Detection System**
  Published in the [**International Journal of Research in Engineering, Science, and Management**](#)
  (April 2020)
  Developed an Intrusion Detection System (IDS) using Random Forest, SVM, and Neural Networks, enhancing cyber threat detection. Optimized model performance and precision-recall metrics, reducing false positives.
- **Agile Project Management** by **Google** - [**Coursera**](#) **Jul. 2024**
- Data Science Methodology by **IBM** – [**Coursera**](#) **Apr. 2024**
- SQL(Advanced) Certification – [**HackerRank**](#) **Mar. 2023**
- **Oracle Cloud** Infrastructure Foundations Associate certification from [**Oracle corporation**](#) **Nov. 2021**