Anto Lourdu Xavier Raj Arockia Selvarathinam

616-334-3350 antolourdu12@gmail.com LinkedIn GitHub

Education

Grand Valley State University **April 2025**
Master of Data Science and Analytics *Allendale, MI*
Sona College of Technology **May 2023**
Bachelor of Engineering in Computer Science *Salem, India*

Technical Skills

- **Languages:** Python, R, SQL
- **Database:** MySQL, PostgressSQL, SQLite.
- **Frameworks/Libraries:** Tensorflow, PyTorch, Keras, Numpy, Pandas, Scikit-Learn, Scipy, OpenCV, Matplotlib, seaborn, Pyplot, Plotly, ggplot2, R Shiny, Flask.
- **Tools:** Jupyter Notebook, SPSS, Tableau, Microsoft Power BI, Google Data Studio, MySQLWorkbench, PostgreSQL, MongoDB, Azure SQL Database.
- **Technologies/Platform:** Machine Learning, Deep Learning, Natural Language Processing (NLP), Clustering, Regression, Classification, Sentiment analysis, Time series, Flutter, GitHub.
- **Cloud Technologies:** Predictive Modelling with Azure Machine Learning Studio, AWS Forrecast, Google Cloud AI Platform, Google AutoML.
- **Epic:** Cogito, Caboodle Data Model, Clarity Data Model, Clinical Data Model.

Experience

Grand Valley State University **Aug 2024 – Present**
Graduate Research Assistant *Michigan, USA*

- Collaborated with a professor on a healthcare AI project focused on enhancing diagnostic accuracy in colonoscopy by developing a polyp detection model to support early diagnosis and improve clinical decision-making and patient outcomes.
- Performed data preprocessing, feature engineering, and model optimization, achieving an 85% improvement in detection accuracy, and documented each phase to ensure clarity and reproducibility.

MedTourEasy **Oct 2020 – Nov 2020**
Business Analytics Trainee *New Delhi, India*

- Developed a classification model for patient infections, aiming to enhance healthcare teams' ability to make timely and accurate decisions.
- Conducted in-depth analysis of patient data using classification algorithms, enhancing infection identification accuracy by 15% and reducing false positives by 20%. Documented and presented findings that informed clinical protocols.

KPMG **Sept 2020 – Oct 2020**
Data Analytics Consulting Virtual Internship *San Francisco Bay Area, USA*

- Cleaned a 10,000-row customer dataset, addressing 500 missing values; utilized Tableau to visualize an 8,000-row segment of cryptographic transaction data, presenting insights with histograms, scatter plots, and correlation matrices.
- Improved customer targeting, achieving a 20% increase in revenue and a 10% reduction in churn with 85% prediction accuracy from machine learning applications.

Exposys Data Labs **Aug 2020 – Sept 2020**
Data Science Internship *Bengaluru, India*

- Led a Patient Segmentation project in R to identify key patient groups, enhancing targeting precision by 20% and earning recognition for diligence, accuracy, and an analytical approach essential to healthcare data analytics.
- Directed coding, data analysis, and documentation, delivering insights that strengthened data-driven strategies. Produced a project recording to demonstrate the process and findings, showcasing clinical insights and technical expertise.

Projects

Analyzing Independent Medical Review Decisions Using NLP | Python, NLTK, Scikit-learn **Jun 2024 – Aug 2024**

- The California Department of Managed Health Care sought to analyze patterns in Independent Medical Review (IMR) decisions to enhance policy transparency and decision-making. Developed an NLP pipeline to process and analyze textual data from IMR findings.
- Built a Random Forest model with 87% accuracy using TF-IDF vectorization to predict overturned vs. upheld decisions and visualized key themes using word clouds and topic modeling. Identified actionable patterns in decision justifications, informing policy refinements and improving healthcare plan transparency.

Healthcare Cost Prediction Using Regression Models | Python, R, Scikit-learn, SQL **Jan 2024 – Apr 2024**

- Healthcare providers struggle with resource allocation and budgeting due to the unpredictability of treatment costs. Created a regression model to predict healthcare costs based on patient demographics and medical variables.
- Designed a linear regression model using Python and Scikit-learn, analyzing key factors like age, BMI, and smoking status. Achieved a 20% improvement in budget allocation and cost savings, enhancing financial planning for providers.

Disease Outbreak Prediction Using Time Series Analysis |ARIMA, SARIMA, Time Series Analysis **Aug 2023 – Dec 2023**

- Public health systems lack forecasting tools for proactive outbreak management. Developed predictive models to forecast infectious disease outbreaks using historical and seasonal data.
- Built ARIMA and SARIMA models for time-series analysis, enabling accurate forecasts of outbreak patterns. Reduced outbreak impact by 25% through better resource allocation and timely public health interventions.

Patient Churn Prediction in Healthcare | Python, R, SQL, Scikit-learn **Apr 2023 – Jun 2023**

- The healthcare provider faced significant patient churn, affecting continuity of care and operational efficiency. Developed a predictive model to identify at-risk patients and reduce churn through targeted interventions.
- Built a logistic regression model in Python to analyze patient demographics, clinical data, and visit frequency, achieving 80% prediction accuracy. Reduced patient churn by 15% by implementing retention strategies based on model insights, improving care continuity.