

# AMRITA NEOGI

| [amritaneogi@arizona.edu](mailto:amritaneogi@arizona.edu) | (520) 427 1767 | Tucson, Arizona | [About Me](#) | [LinkedIn](#) | [GitHub](#) |

## EDUCATION

<b>The University of Arizona, Tucson, Arizona</b> Master of Science in Data Science	Aug 2022 – Dec 2023 (Expected) GPA: 4 / 4
<b>University of Engineering and Management, Jaipur, India</b> Bachelor of Technology in Electrical Engineering	Aug 2013 – May 2017 GPA: 7.66 / 10 (equivalent to magna cum laude)

## SKILLS

- Languages:** R (dplyr, tidyverse, tidymodels, caret, glmnet, Hmisc), Python (Pandas, NumPy, PyTorch, Tensor Flow, Scikit Learn, Matplotlib, tqdm, seaborn), SQL/NoSQL(Joins, correlated subqueries, CTE's)
- Tools & Platforms:** Visualization Tools (Tableau, Looker), Jupyter Notebook, MySQL, Postgres, Teradata, Version Control System (Git), Informatica Power Center, Agile tool( Jira), AWS Services(Athena), Kafka, Google Cloud Storage(GCP), Amazon MTurk, Data Security Platform (Protegrity)
- Data Science and Analytic Skills:** Data Wrangling, Data Engineering, Exploratory Data Analysis (EDA), Machine Learning Algorithms (Classification, Regression, Clustering), Data Visualization, Excel & Spreadsheet Analysis, Predictive Modeling, Feature Engineering, RFM (Recency, Frequency, Monetary) Analysis and Customer Segmentation, Strategic Planning & Decision Making, Critical Thinking & Problem Solving.
- Soft Skills:** Time Management, Task Management, Data Management, Leadership

## PROFESSIONAL EXPERIENCE

<b>The University of Arizona (Department of Pediatrics), Tucson, AZ</b> Graduate Research Assistant	Feb 2023 – Present
<ul style="list-style-type: none"><li><b>Data Optimization Pioneer:</b> Led database optimization efforts, ensuring secure and seamless data queries.</li><li><b>Data Pipeline Management:</b><ul style="list-style-type: none"><li>Orchestrated data cleaning and linkage pipelines, maintaining detailed logs for transparent data refinement.</li><li>Designed data pipelines for MariaDB to PostgreSQL OMOP tables with Python, boosting transfer speed by 30% via Amazon Athena.</li></ul></li><li><b>Post-Survey Processing:</b> Streamlined post-survey tasks, reducing processing time by 20% by integrating MTurk tasks and REDCap surveys.</li><li><b>Data-Driven Healthcare Insights:</b> Applied statistical tests and logistic regression to uncover care patterns, demographics, and engagement factors in patient data analysis.</li></ul>	

<b>Tata Consultancy Services, Mumbai, India</b> Systems Engineer	Mar 2018 – Jul 2022
<ul style="list-style-type: none"><li><b>Extract, Transfer and Load (ETL) Expertise:</b><ul style="list-style-type: none"><li>Leveraged Informatica Power Exchange and PowerCenter 9.x for ETL tasks from various sources ( mainframes, flat files, Teradata, and Enterprise Data Warehouse (EDW), etc.), achieving a 25% enhancement in productivity.</li><li>Created ETL mappings, worklets, mapplets, and reusable transformations, reducing deployment time by 25% while ensuring versatile functionality and seamless database migration across environments.</li><li>Established and maintained data pipelines, resulting in a 30% increase in efficiency for business intelligence, reporting, and analytics.</li></ul></li><li><b>Optimization and Cloud Computing:</b><ul style="list-style-type: none"><li>Identified bottlenecks and optimized performance tuning at the source, target, mapping, and session levels for long-running CI/CD jobs.</li><li>Employed cloud computing (AWS) for ETL optimization, resulting in a 50% reduction in processing time and increased scalability.</li><li>Employed ETL tools to streamline data integration workflows, reducing errors by 28% and lowering infrastructure costs by 2.5%.</li><li>Engineered retrofit solution, integrating 2.5 million customers, while increasing code execution speed by 30%.</li><li>Built data balancing jobs for applications using streaming technologies like Kafka, to ensure completeness and failure alert controls.</li></ul></li><li><b>Documentation and Project Tracking:</b><ul style="list-style-type: none"><li>Collaborated with onshore and offshore teams, conducted unit testing, and maintained a project tracker in JIRA, thereby increasing project visibility and coordination efficiency by 20%.</li><li>Oversaw ETL code repositories with continuous integration practices, resulting in a 15% reduction in defects and improved development cycles.</li><li>Presented findings and insights in a clear and influential manner, driving informed decisions and project success.</li></ul></li><li><b>Leadership and Cross-Functional Expertise:</b><ul style="list-style-type: none"><li>Led a 12-member ETL development team, focusing on data transformation precision, quality, and performance enhancements.</li><li>Guided junior developers in data modeling best practices and fundamental data integration, fostering holistic skill development.</li></ul></li></ul>	

## CERTIFICATION

Google Business Intelligence by Coursera	In Progress
<a href="#">Google Project Management by Coursera</a>	July 2023
<a href="#">Google Data Analytics by Coursera</a>	June 2023

## PROJECTS

<ul style="list-style-type: none"><li><a href="#">Identifying Leaf Phenology of Deciduous Broadleaf Forests from PhenoCam Images</a><ul style="list-style-type: none"><li>To develop a tool for predicting leaf phenology in deciduous broadleaf forests across various sites using CNN Regression.</li><li>Three methods to be employed: AlexNet, ResNet-50, ResNet-101.</li></ul></li></ul>	In Progress
<ul style="list-style-type: none"><li><a href="#">Uber Data Analytics</a><ul style="list-style-type: none"><li>Developed a robust and efficient data pipeline and analytics system for Uber data.</li><li>Extracted and transformed data at a rate of 500 records per second using Google Cloud Storage and Mage ETL.</li><li>Boosted data accessibility by 30% with an interactive Looker dashboard translating raw data into actionable visuals.</li></ul></li></ul>	Aug 2023
<ul style="list-style-type: none"><li><a href="#">Credit Card Fraud Detection</a><ul style="list-style-type: none"><li>Developed a fraud detection model to identify and mitigate fraudulent activities effectively.</li><li>Utilized SMOTE Technique to address data imbalance; employed ML algorithms: Decision Tree, Logistic Regression, Random Forest, and Naive Bayes.</li><li>The best model obtained an accuracy of 92% and a precision of 97%.</li></ul></li></ul>	Aug 2023
<ul style="list-style-type: none"><li><a href="#">Salary Prediction</a><ul style="list-style-type: none"><li>Optimized a salary forecasting model by utilizing Gradient Descent to predict income based on years of professional experience.</li><li>Attained a Mean Square Error (MSE) of 6.3% for the model.</li></ul></li></ul>	July 2023
<ul style="list-style-type: none"><li><a href="#">Design and Implementation of an Image Classifier using CNN</a><ul style="list-style-type: none"><li>Evaluated a deep convolutional network (CNN) for extensive image classification and accomplished an accuracy of 91.21%.</li><li>Utilized Python libraries: NumPy, Pandas, and PyTorch.</li></ul></li></ul>	Dec 2022

## SCHOLARSHIP & AWARDS

<ul style="list-style-type: none"><li><b>University of Arizona, Department of Pediatrics</b> 0.5 FTE (Full Time Equivalent) with Complete Tuition Remission</li></ul>	Feb 2023, Aug 2023
<ul style="list-style-type: none"><li><b>Contextual Master Award, Tata Consultancy Services</b> Promotion as Senior ETL Developer along with salary increment of 14% and bonus of 20,000 INR.</li></ul>	Mar 2022
<ul style="list-style-type: none"><li><b>NEN Champions Runners-Up Award</b> For contributions to entrepreneurship education and innovation at the National Entrepreneurship Network by Wadhwani Foundation</li></ul>	Mar 2016