AMRITA NEOGI

| amritaneogi@arizona.edu | (520) 427 1767 | Tucson, Arizona | About Me | LinkedIn | GitHub |

EDUCATION

The University of Arizona, Tucson, Arizona — Aug 2022 – Dec 2023 (Expected)
Master of Science in Data Science — *GPA: 4 out of 4 scale*

University of Engineering and Management, Jaipur, India — Aug 2013 – May 2017
Bachelor of Technology in Electrical Engineering — *GPA: 7.66 out of 10 scale*

SKILLS

- **Languages:** R (dplyr, tidyverse, tidytext, caret, glmnet, Hmisc), Python (Pandas, NumPy, PyTorch, Tensor Flow, Scikit Learn, Matplotlib, tqdm, seaborn), SQL/NoSQL
- **Tools & Platforms:** Visualization Tools(Tableau, Looker), Jupyter Notebook, MySQL, PostgreSQL, Teradata, Version Control System(Git), Informatica, Agile tool(Jira), AWS Services(Athena), Google Cloud Storage(GCP), Amazon MTurk.
- **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, Problem Solving.
- **Soft Skills:** Time Management, Task Management, Data Management, Leadership

PROFESSIONAL EXPERIENCE

The University of Arizona (Department of Pediatrics), Tucson, AZ — Feb 2023 – Present
Graduate Research Assistant

- Optimize database programs for seamless data querying and ensure secure data procedures and transfers for system integrity.
- Execute data cleaning and data linkage pipeline protocols, while meticulously maintaining detailed work logs for replicable data refinement.
- Create Data Pipelines from MariaDB to PostgreSQL OMOP table using Python and utilize Amazon Athena, increasing data transfer speed by 30% compared to previous methods.
- Create MTurk tasks, integrate REDCap surveys, and analyze data using R upon survey completion, reducing post-survey processing time by 20%.
- Analyze patient data via statistical tests and logistic regression to uncover care patterns, demographics, and factors impacting engagement and adherence.

Tata Consultancy Services, Mumbai, India — Mar 2018 – Jul 2022
Systems Engineer

- **Extract, Transfer and Load (ETL) experience:**
  - Utilized Informatica Power Exchange and PowerCenter 9.x for data extraction, transformation, and loading from various sources, including mainframes, flat files, Teradata, and Enterprise Data Warehouse (EDW).
  - Crafted ETL mappings, worklets, mapplets, and reusable transformations with diverse functionalities and migrate database objects across different environments with a 25% reduction in deployment time and improved efficiency.
  - Established and sustained data pipelines, resulting in a 30% increase in efficiency for business intelligence, reporting, and analytics purposes.
- **Optimization and Cloud Computing:**
  - Identified bottlenecks and optimized performance tuning at the source, target, mapping, and session levels for long-running CI/CD jobs.
  - Employed cloud computing (AWS) for ETL optimization, resulting in a 50% reduction in processing time and increased scalability.
  - Utilized ETL tools to streamline data integration workflows, minimizing errors by 28%, enhancing data quality, while decreasing 2.5% of the infrastructure costs.
  - Engineered a retrofit solution leveraging contextual knowledge, resulting in the integration of 2.5 million customers for the client, and concurrently boosting performance and code execution speed by 30%.
- **Documentation and Project Tracking:**
  - Collaborated with onshore and offshore data stewards and application development leads to maintain project tracker using JIRA, resulting in a 20% increase in project visibility and coordination efficiency.
  - Oversaw ETL code repositories through continuous integration practices, conducting code reviews to uphold code quality and standards, leading to a 15% decrease in defects and improved development cycle.
  - Communicated findings and insights to stakeholders in a lucid and influential fashion, fostering well informed decisions and driving project success.
- **Leadership and Cross-Functional Expertise:**
  - Led a 12-member ETL development team, driving the design and implementation of data transformation processes for precise data integration into the warehouse, while prioritizing data quality and performance enhancements.
  - Guided junior developers on data modeling best practices and fundamental data integration, demonstrating a holistic skill development approach.

CERTIFICATION

- Google Business Intelligence by Coursera — In Progress
- Google Project Management by Coursera — July 2023
- Google Data Analytics by Coursera — June 2023

PROJECTS

- Identifying Leaf Phenology of Deciduous Broadleaf Forests from PhenoCam Images — In Progress
  - Project goal: Develop a tool for predicting leaf phenology in deciduous broadleaf forests across various sites using CNN Regression.
  - Three methods to be employed: AlexNet, ResNet-50, ResNet-101.
- Uber Data Analytics — Aug 2023
  - Project goal: Execute an Uber data project to showcase data expertise, delivering actionable insights and enhanced data accessibility.
  - Extracted and transformed data at a rate of 500 records per second using Google Cloud Storage and Mage ETL
  - Boosted data accessibility by 30% with an interactive Looker dashboard translating raw data into actionable visuals.
- Credit Card Fraud Detection — Aug 2023
  - Project goal: Build robust fraud detection models for credit card transactions.
  - Utilized SMOTE Technique to address data imbalance.
  - Employed ML algorithms: Decision Tree, Logistic Regression, Random Forest, and Naive Bayes.
  - The best model obtained an accuracy of 92% and a precision of 97%
- Salary Prediction — July 2023
  - Project goal: Develop an optimized model using Gradient Descent to forecast salary considering years of experience.
  - Attained a Mean Square Error (MSE) of 6.3% for the model.
- Design and Implementation of an Image Classifier using CNN — Dec 2022
  - Project goal: Evaluate a deep convolutional network(CNN) for extensive image classification.
  - Accomplished an accuracy of 91.21%.
  - Utilized Python libraries: NumPy, Pandas, and PyTorch.

SCHOLARSHIP & AWARDS

- **University of Arizona, Department of Pediatrics** — Feb 2023, Aug 2023
  0.5 FTE (Full Time Equivalent) with Complete Tuition Remission
- **Contextual Master Award, Tata Consultancy Services** — Mar 2022
  Promotion as Senior ETL Developer along with salary increment of 14% and bonus of 20,000 INR.
- **NEN Champions Runners-Up Award** — Mar 2016
  For contributions to entrepreneurship education and innovation at the National Entrepreneurship Network by Wadhwani Foundation