AMRITA NEOGI

| amritaneogi@arizona.edu | (520) 427 1767 | Tucson, Arizona | LinkedIn | About Me | 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), SQL, Java
- **Tools & Platforms:** Tableau, RStudio, Jupyter Notebook, MySQL, PostgreSQL, Teradata, Version Control System (Git), Informatica, Agile tool (Jira)
- **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, 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
- Thoroughly extract data from source into standardized formats and build analytical project databases in REDCap.
- Skillfully optimize database programs for seamless data querying and ensure secure data procedures and transfers for system integrity.
- Precisely execute data cleaning and data linkage pipeline protocols, while meticulously maintaining detailed work logs for replicable data refinement.
- Skillfully create Data Pipelines from MariaDB to PostgreSQL OMOP table using Python scripts.
- Efficiently create MTurk tasks, integrate REDCap surveys, and seamlessly extract and analyze data using R upon survey completion.
- Skillfully analyze patient data via statistical tests and 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:**
  - Skillfully adept in utilizing 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).
  - Expertly crafted ETL mappings, worklets, mapplets, and reusable transformations with diverse functionalities and migrate database objects across different environments.
  - Effectively established and sustained data pipelines to facilitate the integration and processing of data for the purposes of business intelligence, reporting, and analytics.
- **Optimization and Cloud Computing:**
  - Thoroughly identified bottlenecks and optimized performance tuning at the source, target, mapping, and session levels for long-running CI/CD jobs.
  - Effectively employed cloud computing (AWS) for ETL optimization, resulting in a 50% reduction in processing time and increased scalability.
  - Proficiently utilized ETL tools to streamline data integration workflows, minimizing errors by 28%, enhancing data quality, while decreasing 2.5% of the infrastructure costs
- **Documentation and Project Tracking:**
  - Closely collaborated with onshore and offshore data stewards and application development leads to maintain project tracker using JIRA.
  - Diligently oversaw ETL code repositories through continuous integration practices, conducting code reviews to uphold code quality and standards.
  - Effectively conveyed findings and insights to stakeholders in a lucid and influential fashion.
- **Leadership and Cross-Functional Expertise:**
  - Diligently mentored 12 junior developers in ETL job development best practices with Informatica, cultivating a collaborative atmosphere, and openly communicated project status.
  - Effectively guided junior developers on data modeling best practices and fundamental data integration, demonstrating a holistic skill development approach.

CERTIFICATION

IBM Data Science by Coursera	**In Progress**
Google Data Analytics by Coursera	**June 2023**
Google Project Management by Coursera	**July 2023**

PROJECTS

- **Identifying Leaf Phenology of Deciduous Broadleaf Forests from PhenoCam Images**   **In Progress**
  The project is currently in progress. The main aim of this study is to develop a tool which can predict leaf phenology in deciduous broadleaf forest across multiple sites, using CNN Regression. 3 methods to be used including - AlexNet, RestNet-50, RestNet-101
- **Credit Card Fraud Detection**   **Aug 2023**
  The project aims to build a fraud detection model for credit card transactions. SMOTE Technique is applied to manage data imbalance.
  ML algorithms include Decision Tree, Logistic Regression, SVM, and Naive Bayes. The best model achieved an accuracy of XX%
- **Salary Prediction**   **July 2023**
  This project aims to create an optimized model using Gradient Descent, to predict the salary based on the years of experience.
  The model reduced the MSE (Mean Square Error) from 91.2% to only 6.3%.
- **Design and Implementation of an Image Classifier using CNN**   **Dec 2022**
  The project entails evaluation of a deep convolution network for largescale image classification and achieved an accuracy of 91.21%, utilizing Python libraries like NumPy, Pandas, and PyTorch

SCHOLARSHIP & AWARDS

- **University of Arizona, Department of Pediatrics**   **Feb 2023, Aug 2023**
  0.5 FTE with Full 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