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 / 4*
University of Engineering and Management, Jaipur, India Aug 2013 – May 2017
Bachelor of Technology in Electrical Engineering *GPA: 7.66 / 10 (equivalent to magna cum laude)*

SKILLS

- **Languages:** R (dplyr, tidyverse, tidytext, caret, glmnet, Hmisc), Python (Pandas, NumPy, PyTorch, TensorFlow, 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 Platform(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

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

- **Data Optimization Pioneer**: Led database optimization efforts, ensuring secure and seamless data queries.
- **Data Engineering and Pipeline Optimization**:
  - Orchestrated data cleaning and linkage pipelines, maintaining detailed logs for transparent data refinement.
  - Designed data pipelines for MariaDB to PostgreSQL OMOP tables with Python, boosting transfer speed by 30% via Amazon Athena.
- **Post-Survey Processing**: Streamlined post-survey tasks, reducing processing time by 20% by integrating MTurk tasks and REDCap surveys.
- **Data Analytics**: Applied statistical tests and logistic regression to uncover care patterns, demographics, and engagement factors in patient data.

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

- **Extract, Transfer and Load (ETL) Expertise:**
  - 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.
  - Created ETL mappings, worklets, mapplets, and reusable transformations, reducing deployment time by 25% while ensuring versatile functionality and seamless database migration across environments.
  - Established and maintained data pipelines, resulting in a 30% increase in efficiency for business intelligence, reporting, and analytics.
  - Created UNIX shell scripts and tailored PLSQL scripts as required for the purpose of conducting validation tests.
- **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.
  - Employed ETL tools to streamline data integration workflows, reducing errors by 28% and lowering infrastructure costs by 2.5%.
  - Engineered retrofit solution, integrating 2.5 million customers, while increasing code execution speed by 30%.
  - Built data balancing jobs for applications using streaming technologies like Kafka, to ensure completeness and failure alert controls.
- **SDLC Involvement and Project Management:**
  - Engaged in all phases of the SDLC, from initial analysis and requirements gathering alongside Data Science and Investment Science teams, to design, development, testing, and successful software deployment.
  - 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%.
  - Oversaw ETL code repositories with continuous integration practices, resulting in a 15% reduction in defects and improved development cycles.
  - Presented findings and insights in a clear and influential manner, driving informed decisions and project success.
- **Leadership and Cross-Functional Expertise:**
  - Led a 12-member ETL development team, focusing on data transformation precision, quality, and performance enhancements.
  - Guided junior developers in data modeling best practices and fundamental data integration, fostering holistic skill development.

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
  - To develop a tool for predicting leaf phenology in deciduous broadleaf forests across various sites using deep learning (CNN Regression).
  - Three methods to be employed: AlexNet, ResNet-50, ResNet-101.
- Uber Data Analytics Aug 2023
  - Developed a robust and efficient data pipeline and analytics system for Uber data.
  - 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
  - Developed a fraud detection model to identify and mitigate fraudulent activities effectively.
  - 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
  - Optimized a salary forecasting model by utilizing Gradient Descent to predict income based on years of professional experience.
  - Attained a Mean Square Error (MSE) of 6.3% for the model.
- Design and Implementation of an Image Classifier using CNN Dec 2022
  - Evaluated a deep convolutional network (CNN) for extensive image classification and 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