

LABANNYA DEBNATH

(201) 908-1433 | ldebnath@saintpeters.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

Saint Peter's University

Expected May 2028

B.S. in Data Science and Artificial Intelligence, Minor in Statistics

GPA: 4.00

Dean's List - Recognized for Academic Excellence, 2025

Coursework: Statistical Computing with R, Data Mining, Quantitative Analysis, Decision Support Systems

SKILLS

Programming & Analysis: Python (Pandas, NumPy, scikit-learn), R, SQL (Joins, CTEs, Window Functions, Aggregations), Data Analysis, Statistical Analysis

AI, Machine Learning & Data Engineering: Machine Learning, Deep Learning (CNNs), Clustering, ETL Pipelines, Data Modeling, Data Warehousing (Star Schema), RAG, LLMs, Vector Search (FAISS, Chroma, LlamaIndex)

Visualization & Business Intelligence: Power BI, Tableau, Business Intelligence, Data Visualization, Microsoft Excel, Google Sheets, dbdiagram.io

Platforms & Tools: Git/GitHub, PostgreSQL, SQLite, Streamlit, Gradio

Core Strengths: Time Management, Team Collaboration, Communication & Presentation Skills

EXPERIENCE

AI Engineering Extern

Aug 2026 – Oct 2026

Pfizer

- Built a Python pipeline using OCR engines and PDF parsing libraries to extract and structure key data fields from scanned and digital clinical supply documents into JSON with field-level coordinates.
- Implemented a RAG system using vector search (FAISS, Chroma, LlamaIndex) and LLMs to enable intelligent semantic search and querying across unstructured document sets.
- Developed classification and routing logic to categorize documents by type, and deployed a searchable Streamlit/Gradio chatbot interface for querying extracted data, reducing manual document review effort.

Research Assistant

May 2025 – Aug 2025

Saint Peter's University

- Collected, cleaned, and organized MRI imaging datasets for a faculty-led research applying CNNs to cancer cell detection, using Python and R for data preprocessing and management.
- Conducted literature reviews on deep learning approaches to medical image classification and synthesized findings to inform model architecture and research direction.
- Performed exploratory data analysis and basic statistical evaluation in Python and Excel to support dataset validation ahead of neural network training.

PROJECTS

Movie Rental Transactions Data Warehouse (SQLite) | [Link](#)

- Designed a star-schema data warehouse in SQLite for 500+ rental transactions, modeling fact and dimension tables with primary and foreign keys to ensure referential integrity and scalable querying.
- Built SQL analytics using multi-table joins, CTEs, window functions, and aggregations to analyze revenue trends, customer behavior, genre performance, and store KPIs.
- Implemented data quality checks for nulls, duplicates, orphaned keys, and invalid records, and optimized queries through indexing and analysis.

AI Against Climate Gentrification (Python, scikit-learn) | [Link](#)

- Built an ETL pipeline in Python and pandas to clean, standardize, and merge multi-source datasets across county, state, and year keys into a unified analytical dataset.
- Engineered year-over-year AQI and housing price change features and computed correlation analysis to evaluate the relationship between air quality improvement and rising home prices, a proxy for climate gentrification risk.
- Applied K-means clustering with feature scaling to segment counties into risk profiles, visualized with Seaborn scatterplots and cluster centroids to support policy insights.

Library Circulation Database (SQLite) | [Link](#)

- Designed a normalized relational schema for a library system across 8 entities (authors, books, patrons, branches, copies, loans, fines) with primary and foreign key constraints and cascading deletes to enforce data integrity.
- Built reporting queries and views to analyze checkout trends, overdue loan rates, fine payment rates, and patron activity across the library system.
- Implemented data quality checks to catch null keys, duplicate records, and orphaned foreign keys, and optimized query performance through indexing, validated with EXPLAIN QUERY PLAN benchmarking.