Donnoban Maldonado

Long Island, NY | mdonnoban@gmail.com | https://www.linkedin.com/in/donnoban-maldonado | github.com/donnmaldonado

EDUCATION

State University of New York at Old Westbury   May 2025
Bachelor of Science in Computer Science
Honors: Summa Cum Laude (GPA: 3.98)
Relevant Coursework: Machine Learning, Data Mining, Database Management Systems, Business Analytics

SKILLS

Languages: Proficient: Python (4yrs) · SQL (3yrs) · Intermediate: Java (2yrs) · JavaScript(2yrs) · Beginner: C++(1yr)
Technologies: AWS · GCP · Git · Django · Pandas · Numpy · Scikit-Learn · Tableau · PowerBI · Excel · Sheets
Bilingual: English · Spanish

EXPERIENCE

Data Science Research Assistant   January 2025 - Present
SUNY RESEARCH FOUNDATION   *Old Westbury, NY*

- Designed and implemented an **ETL pipeline** to collect and structure Discord server data for analysis and machine learning, increasing the training dataset by 43% using Python, Pandas, and the Google Sheets API.
- Led deployment and integration of the above-mentioned system across 3 partner institutions using **Google Cloud Platform**, collaborating with researchers to optimize performance and ensure data compliance.
- **Uncovered patterns** in transfer student challenges by analyzing 23,000+ Reddit posts with a mixed-methods NLP pipeline, showing that **LLMs outperform traditional models** in extracting actionable insights from unstructured text data.
- Presented work on data pipeline design and ML preparation at the **SUNY Undergraduate Research Conference (SURC)**, SUNY Binghamton.

PROJECTS

MagicSheets | Github   April 2025

- Built a **generative AI web app** using Django, Python, Bootstrap, PostgreSQL, and ChatGPT-4o to create customizable educational worksheets for teachers, parents, and tutors.
- Implemented features for multiple worksheet types, grade-level filtering, answer key generation, and user-driven regeneration tools (e.g., "make it more descriptive").
- Integrated community functionality for sharing, rating, and versioning worksheets to encourage collaboration and iterative improvement.
- Deployed on Render with a structured PostgreSQL backend, enabling reliable performance and **easy data extraction** for research and analytics.

IMDb Review Sentiment Analysis | Github   December 2024

- Preprocessed and vectorized **50,000 IMDb reviews** using Python and Word2Vec, cleaning unstructured text by removing HTML, emojis, URLs, and duplicates.
- Trained and tuned multiple models (**Logistic Regression, SVM, XGBoost, MLP**), achieving an **89% F1-score** with MLP and selecting Logistic Regression for its balance of accuracy and interpretability.
- Performed **EDA** and word cloud analysis to identify sentiment patterns and refine preprocessing steps.
- Improved model performance by selecting optimal embeddings (500-dim) when model hyperparameter tuning yielding minimal gains over well-generalized default models.

CERTIFICATIONS

Microsoft Certified: Power BI Data Analyst Associate   January 2025
Preparing data; modeling data; visualizing and analyzing data; and deploying and maintaining assets

IBM Data Science Professional   January 2023
Data science and machine learning techniques

AWS Cloud Practitioner   July 2023
Cloud fluency and foundational AWS knowledge