

MARK LOUISE BADUA

BS Physics graduate from University of the Philippines Diliman, Looking for opportunities to be involved in the field of Startups, Technology, Data Science, Artificial Intelligence, and Sustainable Development.

✉ markbadua92@gmail.com www.linkedin.com/in/mark-louise-badua-140854184

Education

University of the Philippines Diliman

Quezon City, Metro Manila

Bachelors of Science in Physics

Aug 2018 - July 2023

- **Undergraduate Coursework:** Python, C, Data Analysis, Physical Modeling, Statistics, Calculus, Differential Equations, Linear Algebra, Signal/Image Processing, Classification, Dimensionality Reduction.
- **Graduate Coursework:** Introduction to High Performance Computing using Julia, Natural Language Processing, Large Language Models, Diffusion Models, Hyperspectral Unmixing.
- **Awards:** University Scholar (1st Sem AY 2020-2021, 2nd Sem AY 2021-2022), College Scholar (2nd Sem AY 2020-2021, 1st Sem AY 2021-2022)

Relevant Experience

Certchain

Remote

Junior AI Developer

Jan 2024 - Aug 2024

- Migrated data lake and endpoints from Azure to AWS.
- Created and maintained multiple in-house dashboards using Streamlit to monitor data operations and quality.
- Designed and implemented the backend of a self-service Retrieval-Augmented Generation (RAG) pipeline.
- Developed an in-house library to automatically test the quality of chatbot responses, ensuring reliability and performance.
- Designed and implemented a scheduled-based ETL pipeline
- Migrated from a scheduled-based ETL pipeline to an event-driven architecture, improving efficiency and scalability.
- Tools: Python · Azure Blob Storage · Azure OpenAI · LangChain · FastAPI · Streamlit · Git · Docker · Bash · SQL · AWS S3 · AWS Batch · AWS CodeBuild · AWS Lambda · AWS RDS · AWS Glue · AWS Step Functions · AWS Athena · MongoDB Atlas Search · Weaviate · Chroma DB

Captivate

Remote

AI Engineer

July 2024 - Aug 2024

- Collaborated with TD Synnex through Captivate to initiate and lead the development of a Retrieval-Augmented Generation (RAG) pipeline for a chatbot project
- Initiated the design and implementation of the backend of a self-service RAG pipeline, collaborating with team members to bring it to completion.
- Tools: Python · LangChain · FastAPI · Docker · Bash · GitHub · Azure Blob Storage · Azure CosmosDB · Azure AI Search · Azure OpenAI

Nidec Drives

Remote

Data Scientist

March 2024- August 2024

- Developed a product-comparison chatbot as a Proof of Concept (POC).
- Developed a sales coach chatbot as a Proof of Concept (POC).
- Performed exploratory data analysis (EDA) to uncover patterns, identify anomalies, and gain insights from datasets using various statistical and visualization tools.
- Solely developed and implemented all features of a mobile app using Flutter.
- Tools: Python · NumPy · Pandas · Dart · SQL · Flutter · Streamlit · GitHub · Jupyter Notebooks · OpenAI API · Bash · Supabase · Azure Container Registry · Azure App Service

Control Techniques

Hybrid

Data Science Intern

Nov 2023 - March 2024

- Led a team of 8 data science interns to prototype a web application designed for use by salespeople, overseeing all stages from initial concept to functional prototype.
- Developed a geospatial algorithm that recommends clients for salespeople to visit based on their starting location and client preferences, optimizing travel routes.
- Tools: Python · NumPy · Pandas · SQL · Streamlit · GitHub · Jupyter Notebooks · Bash · Azure Container Registry · Azure App Service

Omdena

Remote

Junior Machine Learning Engineer (Volunteer work)

Nov 2023 - Dec 2023

- Collaborated with 50 data science professionals and volunteers worldwide from 20+ countries in collaboration with Voy Finance to develop: Greenwashing Detection Model, ESG Scoring Model, Automated Text Insight Company Reports Generation.
- Researched building a custom entity linking model to identify companies from text.
- Developed Python script for Custom Entity Linking, utilizing user-defined paths for knowledge base and trained entity linking model
- Tools: SQLite · Python · Bash · spaCy · SQL · Git

Instrumentation Physics Laboratory

Quezon City, Metro Manila

Air Quality Data Analyst/Researcher

Sep 2020 - July 2023

- Used Time-Resolved Statistics and Discrete Fourier Transforms to understand the temporal behavior of ambient particulate matter at the Lung Center of the Philippines.
- Reviewed and presented papers related to air quality.
- Skills: Statistics · Data Integration · Data Visualization · Data Cleaning · Git · Python · Bash

Publication

Pre-ECQ versus ECQ: a comparison of PM1 measurements

Legazpi City, Albay

First author

Oct 19 2022

- Coauthored a paper and presented at the 40th SPP Physics Conference that analyzed and compared ambient air quality levels before and during the pandemic.
- Utilized Time-Resolved Statistics and Discrete Fourier Transforms to find seasonalities and correlations related to anthropogenic particulate matter.
- Visualized findings by utilizing boxplots and histograms (1D and 2D, in Cartesian and Polar).
- Automated the consolidation of previously scattered measurements, originally spread across various directories and tar.gz files, into a single, easily accessible CSV file.
- Enhanced time series data quality through a data cleaning process that employed rate of change analysis to discern and eliminate local maxima deemed unnecessary.
- Tools: Python (NumPy, Pandas, Matplotlib, Seaborn, scikit-learn), Git, LaTeX, Bash.

Relevant Undergraduate and Graduate Programming Coursework

Physics 312 - Advanced Mathematical Physics II

Quezon City, Metro Manila

Feb 2023 - July 2023

- Enrolled in a PhD course that covered the following topics: Large Language Models (LLM), Diffusion Models, and Hyperspectral Unmixing.
- Suggested a modification to Transformers architecture tailored for classifying physics articles.
- Tools: Python, Jupyter Notebook, Google Colab, Bash

Physics 215 - Computation Methods of Physics

Quezon City, Metro Manila

Sep 2022 - Jan 2023

- Completed an advanced Master's-level course focused on high-performance computing utilizing the Julia programming language
- Explored the differences and advantages of Julia over other programming languages such as Python and C
- Tools: Julia

CS 11 - Computer Programming I

Sep 2022 - Jan 2023

- Developed proficiency in fundamental programming constructs, such as data types, control structures, and functions, by solving coding problems using the C programming language.
- Learned and implemented various data structures, including arrays, linked lists, stacks, and queues.
- Tools: C, Visual Studio Code.

App Physics 157 - Computational Analysis and Modeling in Physics

Feb 2022 - May 2022

- Explored Monte Carlo Methods and Complex Network Tools: Diffusion Limited Aggregation, Ising Model, Cellular Automata, Agent-based models, Graph Models, Network Properties.
- Explored Image and Video Processing Methods: Digital Image Formation and Enhancement, Feature Extraction from images and video, Fourier Transform applications in Image Processing.
- Explored Machine Learning: Dimensionality Reduction (Principal Component Analysis), Regression (Linear Regression; Overfitting, Underfitting, and Cross-Validation), Classification (Decision Trees).

- Classified Galaxies using Decision Trees and analyzed SDSS Galaxy Spectra using Principal Component Analysis.
- Tools: Python (Numpy, Matplotlib, scikit-learn), Google Colab.

App Physics 155 - Computer Methods in Physics I

Sep 2020 - Dec 2020

- Acquired proficiency in basic programming constructs, including variables, data types, operators, conditional statements, loops, functions, arrays/lists, and control flow, with a focus on Python syntax.
- Computationally solved Derivatives, Integrals, Linear equations, Nonlinear Equations, Ordinary and Partial Differential Equations, and Fourier Transforms.
- Tools: Python (Numpy, Matplotlib, scikit-learn), Google Colab.

Technical Skills

Programming Languages: Python, Bash, Julia, C, Dart

Databases: SQL (Supabase, PostgreSQL, MySQL), NoSQL (MongoDB Atlas), Vector Databases (ChromaDB, Weaviate, MongoDB Atlas, CosmosDB, Azure AI Search)

Frameworks and Libraries: LangChain, Streamlit, Flutter

Collaboration: Git (Github), Google Colab.