

# John Paul M. Curada

(+63) 919 328 0443 / johncurada.02@gmail.com / Quezon City, Metro Manila

[linkedin.com/in/jpcurada](https://www.linkedin.com/in/jpcurada) / [datacamp.com/portfolio/jpcurada](https://datacamp.com/portfolio/jpcurada) / [github.com/JpCurada](https://github.com/JpCurada) / [streamlit.io/user/jpcurada](https://streamlit.io/user/jpcurada)

---

Computer Science student at Polytechnic University of the Philippines with proven AI/ML expertise through 5+ national competition wins and international recognition. Delivered production ML systems processing 7B+ records at Frontier Tower Associates. Specialized in Data Science, NLP, computer vision, and agentic AI systems with hands-on experience in TensorFlow, PyTorch, and cloud deployment (AWS, GCP). Passionate about developing scalable Data and AI solutions for real-world impact.

**Skills/Technologies:** TensorFlow, Keras, PyTorch, Scikit-learn, YOLOv8, Ultralytics, ONNX, Hugging Face Transformers, Knowledge Distillation, Model Quantization, Google Agentic Development Kit, Flowise, n8n, LangChain, Langgraph, Pydantic AI, OpenAI API, Multi-agent Systems, Python, SQL, BigQuery, Pandas, NumPy, Data Pipelines, ETL, Feature Engineering, GCP, Cloud Run, AWS, Docker, Model Deployment, CI/CD, Power BI, Looker, Plotly, Matplotlib, Seaborn, PyGWalker, Dash, Git, GitHub, GitLab, React, TypeScript, JavaScript, Java, C, FastAPI, Flask, Django, Streamlit, Supabase, PostgreSQL, MySQL

---

## Experience

---

### IT - Data and Artificial Intelligence Intern

Frontier Tower Associates Philippines, BGC, Taguig

September 2025 – March 2026 | Hybrid

- Built a geospatial data mining pipeline using Computer Vision and Google Earth Engine to extract infrastructure data from satellite maps; utilized Levenshtein distance to ensure high-accuracy data loading into BigQuery.
- **Developed a high-precision OCR platform** powered by **Tesseract.js** and **Docling** to automate and validate data extraction from power billing statements. Implemented a **bounding box annotation** system to target specific document regions, integrating **PyMuPDF** for seamless document processing and LLM-ready data preparation.
- Conducted R&D on multi-agent workflows using **LangGraph**, **Flowise**, and **Pydantic AI** to develop an in-house **Agentic Data Analyst** platform. This tool automates data extraction from Excel, CSV, and BigQuery to generate visualizations and key insights. Successfully deployed the system on **GCP** using **Docker** and **Google Artifact Registry**, integrating **Okta authentication** to ensure secure, employee-only access.

### Freelance AI/ML Python Developer

October 2024 – August 2025

Project-based

- Developed YOLOv8 real-time fabric defect detection system achieving 94% accuracy with 45ms inference time using Ultralytics and ONNX optimization, deployed via Streamlit admin interface with Roboflow integration, reducing manual inspection time by 60%
- Evaluated and implemented transfer learning on TensorFlow Model Garden CNNs for custom fabric dataset using TensorFlow/Keras, improving classification performance by 23% over baseline models
- Built end-to-end IoT fire detection system with Supabase backend, Streamlit interface, and scikit-learn Random Forest classifier achieving 89% detection accuracy for real-time emergency response

### Junior Machine Learning Engineer (Project Based)

Omdena

August 2023 – November 2024 | Remote

Led cross-functional ML teams of 15+ engineers across 4 international projects, delivering production-grade models for healthcare, agriculture, legal tech, and climate analysis domains with measurable community impact

- **Personal Injury Claims Valuation (US):** Conducted comprehensive EDA with boxplot/whisker visualizations for outlier detection and statistical analysis, contributing to ensemble ML-based legal compensation prediction model (Random Forest, XGBoost, Linear Regression) achieving 82% accuracy with RMSE optimization
- **Tuberculosis Analysis Platform (Nigeria):** Led data preparation team, automated Excel-to-unified format processing using Pandas reducing data cleaning time by 80%, developed interactive PyGWalker/Plotly dashboard with time-series analysis deployed via Streamlit Cloud
- **Mango Leaf Disease Detection (Bangladesh):** Spearheaded EDA and built custom CNN dataset loading pipeline using TensorFlow/Keras ImageDataGenerator for agricultural computer vision model, implementing data augmentation (rotation, flip, zoom) and transfer learning (ResNet50, MobileNetV2) enabling real-time disease classification with 91% accuracy
- **Climate Data Analytics (São Paulo, Brazil):** Engineered weather data retrieval system for 645 regions spanning 23 years (2000-2023) using OpenStreet/Weather APIs with custom data manipulation functions (Pandas, NumPy), implementing time-series forecasting (ARIMA, Prophet) processing 15,000+ data points daily

---

## Certifications

---

AWS Certified Cloud Practitioner  
[Credential ID](#)

DataCamp Certified Python Data Associate  
[Credential ID](#)

Datacamp Certified AI Engineer  
[Credential ID](#)

---

## Competitions & Awards

---

- Datacamp Data Analysis and Storytelling Competitions ([Reference](#))
  - **Everyone Can Learn Data Scholarship 2023:** Performed exploratory data analysis on Fossil Records dataset using Python (Pandas, NumPy, Matplotlib, Seaborn) and IMDB database using SQL (joins, aggregations, window functions, subqueries)
  - **Predicting Industrial Machine Downtime (Levels 1-3):** Won all three competition levels by building predictive maintenance classification models using ensemble methods (Random Forest, XGBoost, Gradient Boosting), feature engineering, hyperparameter tuning (GridSearchCV), and imbalanced data handling (SMOTE)
- ASCENT 2023: A Data Storytelling Challenge hosted by UP Limitless Statistical Society.
  - **1st Place** from 100+ participants - Presented COVID-19 epidemiological time-series analysis project using Python (Pandas, Plotly, Matplotlib), implementing moving averages, trend analysis, correlation studies, and interactive dashboards at UP Diliman School of Economics Auditorium
- Agentic AI Hackathon hosted by Gen AI Philippines.
  - **1st Runner Up** from 50+ teams/ Developed multi-agent diabetes management solution using Google Agent Development Kit, React, Firebase, implementing personalized health recommendations, meal planning algorithms, and glucose monitoring workflows
- SIDHI Data Analysis and Visualization Hackathon 2023:
  - Secured **2nd place** by performing data cleaning, analysis, and visualization using Python on a provided complex datasets

---

## Projects

---

### **BisTag: Bisaya-Tagalog Code-Switching Morphological Analyzer**

Built finite-state morphological analyzer performing tokenization, parsing, morphological decomposition, lemmatization, and POS tagging on mixed Bisaya-Tagalog text using Non-Deterministic Finite Automaton for Philippine low-resource language processing.

*Tools: Python, NLP, Finite State Machines, Tokenization, Parsing, Morphological Analysis* | [GitHub](#)

### **Diabeathis: Agentic AI Healthcare Solution**

Developed multi-agent diabetes management platform for Google's APAC Solution Challenge 2025 (1st Runner Up at Agentic AI Hackathon). Architected intelligent agent workflows for personalized health recommendations, meal planning, and glucose monitoring using Google Agent Development Kit.

*Tools: TypeScript, React, Firebase, Google Agent Development Kit, GCP, Multi-agent Systems* | [GitHub](#)

### **SeekLiyab: IoT Fire Detection with Machine Learning**

Engineered IoT fire detection system with Random Forest classifier on sensor fusion data achieving 89% detection accuracy. Implemented real-time anomaly detection and predictive alerting for emergency response coordination.

*Tools: Python, Scikit-learn, Random Forest, IoT, Streamlit, Supabase, Real-time Analytics* | [GitHub](#)

### **Exploralytics: Python Data Visualization Package**

Created and published Python package on PyPI with 15+ Plotly visualization templates for exploratory data analysis, achieving 500+ downloads from data science community.

*Tools: Python, Plotly, PyPI, Data Visualization, Package Development* | [PyPI](#)

### **AWS Cloud Club PUP Skill Builder Platform**

Led serverless gamified learning platform development serving 400+ students using Lambdalith architecture with Docker/ECR, S3, CloudFront, Route53, and SES for automated certificate distribution.

*Tools: Python, AWS Lambda, Docker, ECR, S3, CloudFront, Serverless Architecture* | [Live](#)

---

## Education

---

Polytechnic University of the Philippines

September 2023 - 2027

Manila, Metro Manila, Philippines

Bachelor of Science in Computer Science

Relevant Coursework: *Machine Learning, Natural Language Processing, Computer Vision, Automata Theory, Data Structures & Algorithms, Cloud Computing*

Honorato C. Perez Sr. Memorial Science High School

August 2021 - July 2023

Cabanatuan, Nueva Ecija, Philippines

Valedictorian, Exemplary Leadership Awardee, Best in Mathematics and Science  
Science, Technology, Engineering, and Mathematics Strand