# RALF CEDRIC T. ZABALA

Manila, Philippines ralfcedric30@gmail.com
linkedin.com/in/ralf-cedric-zabala-748901385/

## EDUCATION

**Mapúa University** Manila, Philippines
Bachelor of Science in Computer Science, Specialisation in Data Science September 2026
Magna Cum Laude, Final GWA: 1.4262

## TECHNICAL SKILLS

**Analytics & BI:** SQL, PostgreSQL, Power BI, DAX, Power Query, Microsoft Excel, Pandas
**Data Engineering:** ETL, Dimensional Modelling, Apache Kafka, PySpark, Spark Structured Streaming, Apache Cassandra
**Programming & ML:** Python, scikit-learn, PyTorch, Hugging Face Transformers, K-Means, PCA, NLP
**Tools:** Git, GitHub, FastAPI, Supabase, Azure DevOps, Postman

## PROJECTS

**Retail Sales & Inventory Analytics**
PostgreSQL | SQL | Power BI | DAX | Power Query

- Built an end-to-end PostgreSQL retail analytics solution using raw, staging and analytics layers across eight source tables, profiling and transforming approximately 5,000 orders, 12,680 order-item records and 9,600 inventory snapshots.
- Designed a transaction-product-line star schema and advanced SQL analyses using CTEs, LAG and RANK; developed four Power BI dashboard pages with DAX for sales, product/store, customer and inventory KPIs.

**Customer Segmentation & Market Basket Analysis**
Python | Pandas | RFM | K-Means | PCA | Apriori | FP-Growth | ECLAT

- Engineered RFM features for 4,718 e-commerce customers and fitted four K-Means segments, achieving a 0.596 silhouette score and 0.687 Davies-Bouldin index; visualised customer-segment separation using PCA.
- Built transaction baskets and applied Apriori, FP-Growth and ECLAT with support/confidence/lift analysis to identify candidate product relationships for cross-selling and bundling.

**Real-Time Traffic Data Streaming Pipeline**
Python | Apache Kafka | PySpark | Spark Structured Streaming | Apache Cassandra

- Built an end-to-end real-time traffic pipeline with 10 concurrent Python sensor producers, streaming vehicle-count events through Kafka and Spark Structured Streaming into Cassandra and persisting 6,754 simulated records.
- Implemented JSON schema parsing, null validation, TimeUUID handling, 10-second micro-batches, checkpointing and foreachBatch writes; verified end-to-end ingestion through the Spark UI and CQL queries.

**Insighta - AI-Assisted Insurance Complaint Management Platform**
Next.js | TypeScript | FastAPI | Supabase | DistilBERT

- Developed a complaint-management platform for non-life insurance with public complaint submission/tracking, role-based staff and administrator workflows, analytics and database-backed case management.
- Integrated an NLP service with asynchronous processing and human-review workflows to support complaint categorisation, triage and structured routing.

## PROFESSIONAL EXPERIENCE

**EastWest Bank - Shared Testing Services (STS)** Makati, Philippines
**Quality Assurance Intern - Test Automation** May 2026 - August 2026

- Executed and documented 230+ passed functional, regression, UI and API-oriented test cases across digital-banking workflows, recording expected versus actual outcomes and filing 3+ evidenced defects through Azure DevOps.
- Maintained Katalon automation cases and XPath locators, executed Postman payment and transfer scenarios including InstaPay and PESONet, and investigated failed automation runs; gained exposure to Playwright.

## UNDERGRADUATE RESEARCH

**Adversarially Robust English Smishing Detection**
Python | PyTorch | DistilBERT | Genetic Algorithms | NLP | Adversarial Machine Learning

- Fully designed a hybrid smishing-detection architecture using 10,544 English SMS records, combining cached 768-dimensional frozen DistilBERT embeddings with eight interpretable engineered feature groups whose importance weights were evolved by a genetic algorithm; trained and evaluated seven model configurations.
- Under the adv30 condition, the proposed model achieved 95.45% smishing recall with 36 false negatives versus 117 for the TF-IDF baseline; evaluated the accompanying false-positive trade-off using paired McNemar tests and 10,000-replicate bootstrap confidence intervals. Paper accepted for oral presentation at ICICM 2026.