

Naphat Jakkraphatcharakul

Bangkok, Thailand | +66 93 130 7270 | serenaury.will@gmail.com
linkedin.com/in/serenaury | github.com/Serenaury | serenaury.github.io
Data Science & Analytics co-op candidate | Available Jan–Apr 2027

EDUCATION

Thammasat University

2023 – 2027 (expected)

B.Sc. in Data Science and Innovation (Cybersecurity module), College of Interdisciplinary Studies

GPA 3.33/4.00; last four semesters 3.56

EXPERIENCE

Business Administration Intern, Marketing

Jun 2026 – Aug 2026

UMEDA Co., Ltd. (pharmaceutical company), Bangkok

- Built Excel workbooks that break down pharmaceutical sales by manufacturer and product across government, private and OTC channels, grouped by ATC3 therapeutic class, and ranked both within each channel.
- Compared hospital and reference prices, then projected future prices with Excel trend and forecast tools.
- Turned the rankings and price comparisons into Power BI dashboards and PowerPoint decks for the Marketing team's meetings with executives.

PROJECTS

Thrive: Rehabilitation Data Analysis

Jan 2026 – Present

Academic project for the Ministry of Justice's Kamlangjai program | Python, SQL (DuckDB), Parquet, scikit-learn

- Linked 12 corrections datamart tables with probation data in DuckDB into person-level Parquet tables covering 2M+ people, with a data contract that tags each feature as usable, proxy or protected.
- Compared occupational-grant recipients with non-recipients using direct standardization by sex, age band and prisoner class, applied Benjamini-Hochberg correction across all tested factors, and presented the results in an aggregate-only dashboard.
- Benchmarked multinomial logistic, ordinal and gradient-boosting models against a majority baseline for prisoner class (levels 1–6); found the top features were inputs to the official classification rule, so reported the models as a rule-consistency check, backed by a fairness audit and model card.

CVE-to-ATT&CK Technique Ranking

Sep 2026 – Present

Independent research project | Python, LightGBM (LambdaRank), BM25, MiniLM (sentence-transformers), pandas

- Built a CVE–CWE–CAPEC–CPE knowledge graph that keeps the source and date of every link, and trained a LambdaRank model on 44 text and graph features to rank MITRE ATT&CK techniques for each CVE role.
- Under a pre-registered protocol, reached nDCG@5 0.558 on a sealed test set (158 queries) versus 0.218 for MiniLM embeddings and 0.145 for BM25; 5×5-fold CV ablations traced most of the gain over text-only to platform (CPE) features, as a randomized graph scored the same as the real one.

Student Dropout Risk Prediction

Aug 2025 – Dec 2025

Academic research project | Python, pandas, scikit-learn, XGBoost

- Cleaned course-level records for 20,919 undergraduates (2014–2024), rolled them up to credit-weighted term GPA, and engineered GPA-trend, variability and course-failure features.
- Compared logistic regression, random forest and XGBoost at three checkpoints with a student-level 80/20 split. XGBoost led on ROC-AUC and F1 at every stage: ROC-AUC 0.866 after Term 1, rising to 0.992 after Term 3.

ArguSQL: Desktop SQL Workspace

2026

Personal project | TypeScript, Electron, React, DuckDB

- Built a Windows app that runs SQL directly on local CSV, TSV, Parquet, JSON and JSONL files, with live preview, autocomplete, a visual JOIN builder and CSV export; covered by 185 unit and 96 end-to-end tests.

PUBLICATION

A Machine Learning-Based Student Dropout Risk Prediction and Early Warning System. Asian Health, Science and Technology Reports, 34(4), 2026 (accepted). Second author; presented at AJCC 2026, Sapporo, Japan.

SKILLS

Programming: Python (pandas, scikit-learn, XGBoost), SQL (DuckDB, MySQL), NoSQL (MongoDB), TypeScript

Statistics & Machine Learning: EDA, hypothesis testing, direct standardization, multiple-testing correction, regression, classification, clustering, learning to rank, NLP (text embeddings), model evaluation, fairness audit

Data engineering: data modeling, ETL, data warehousing, data quality checks, data governance, data contracts

BI & analytics: Power BI, Tableau, Excel (pivot tables, forecasting), KPI and dashboard design, pricing analysis

Tools & security: Git, Docker (web project), Linux, digital forensics (Volatility), log analysis, network security

Languages: Thai (native), English (upper intermediate)