PORTFOLIO · 2026

OPEN TO WORK · JAKARTA / REMOTE

# MOHAMMAD ARKANANTA RADITHYA TARATUGANG

## AI / ML ENGINEER · DATA SCIENTIST · SOFTWARE ENGINEER

Information Technology graduate from Institut Teknologi Sepuluh Nopember (ITS) Surabaya specializing in multimodal Deep Learning (IndoBERTweet + Vision Transformer), end-to-end Big Data streaming pipelines (Kafka, Spark, MinIO), and full-stack production systems (FastAPI, Vue 3, Docker).

FIRST AUTHOR · ICAICTA 2026

IELTS ACADEMIC 7.0 · CEFR C1

12+ PROJECTS SHIPPED

GPA 3.62 / 4.00 · S.KOM ITS

Jakarta, Indonesia

+62 822-5477-2727

radithyarkaa@gmail.com

linkedin.com/in/arkananta-radithya

github.com/radithyaarka

kaggle.com/arkaradithya

WhatsApp



OPEN TO WORK ✦ MULTIMODAL AI ✦ BIG DATA PIPELINES ✦ EXPLAINABLE AI ✦ FULL-STACK SHIPPING ✦ ITS GRADUATE ✦ JAKARTA / REMOTE

# CELESTX



SOLO AUTHOR · MULTIMODAL AI RESEARCHER · UNDERGRADUATE THESIS

Depression-signal detection on Indonesian social media: **7,991 tweets** train the text modality, **294 tweet-image pairs** evaluate the fusion. **IndoBERTweet** reads text, a **Vision Transformer** reads images, and the fused signal drives a **3-tier** alert built on **DSM-5** criteria.

**82.10%**
ACCURACY · 294 PAIRS · 5-FOLD CV

**79.10%**
F1-SCORE · 294 PAIRS

**13**
FALSE NEGATIVES · OF 294 CASES

**SUS 80**
USABILITY · EXCELLENT

- **The Problem:** Unimodal text NLP models miss depressive signals on modern social media, where much of the context lives in the images.
- **Multimodal Fusion:** Combines **IndoBERTweet** text features with **Vision Transformer** image features using early fusion + attention at the user level.
- **Dual Persona:** A digital exposure meter for general users, and a digital phenotyping tool (SHAP + attention maps) for clinicians.

PYTORCH · INDOBERTWEET · VISION TRANSFORMER · FASTAPI · REACT.JS · SHAP (XAI) · CHROME EXTENSION · DOCKER

FIRST AUTHOR · ICAICTA 2026 · PRESENTED 20 OCT 2026 · UTAR KAMPAR, MALAYSIA

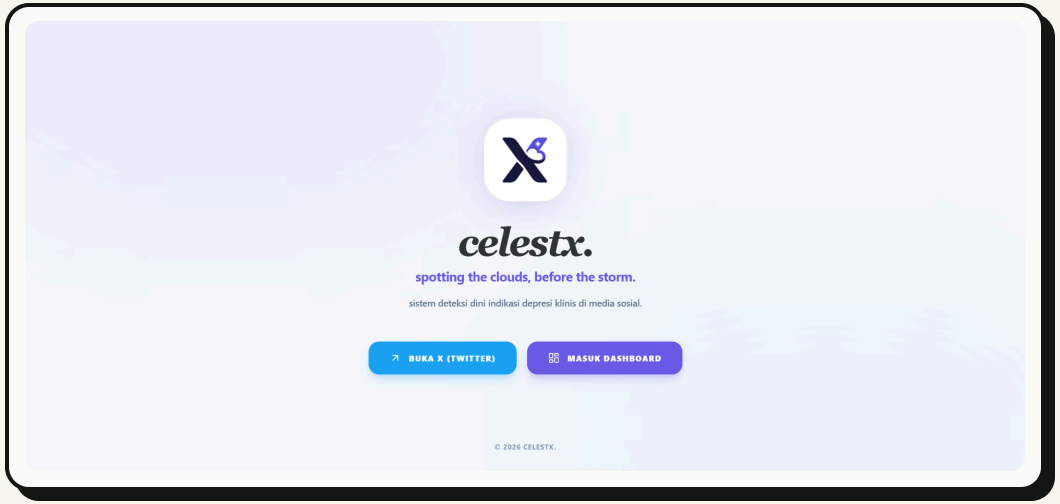



FIG. 2.1 · CELESTX LANDING PAGE

# CELESTX



## PIPELINE & CROSS-MODAL FUSION

- **NLP Layer:** Fine-tuned IndoBERTweet to capture Indonesian social media slang, acronyms, and emotional context.
- **Vision Layer:** Vision Transformer extracting visual features, color palettes, and emotional tone from image feeds.
- **Fusion:** Early fusion + attention that adaptively weighs text and image contribution per user (user-level).

## DUAL PERSONA & EXPLAINABILITY

- **General users:** a digital exposure meter with real-time alerts via the Chrome Extension on Twitter/X.
- **Clinicians:** a clinical dashboard + risk archive, with SHAP values and ViT attention maps explaining predictions.
- **Delivery:** a FastAPI backend and React dashboard connected to the observatory.

**82.10%**
ACCURACY · 294 PAIRS · 5-FOLD CV

**79.10%**
F1-SCORE · 294 PAIRS

**13**
FALSE NEGATIVES · OF 294 CASES

**SUS 80**
USABILITY · EXCELLENT

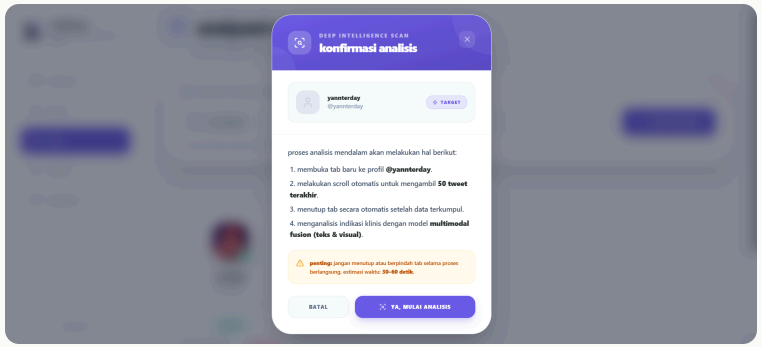



FIG. 3.1 · DEEP SCAN · MODEL PIPELINE

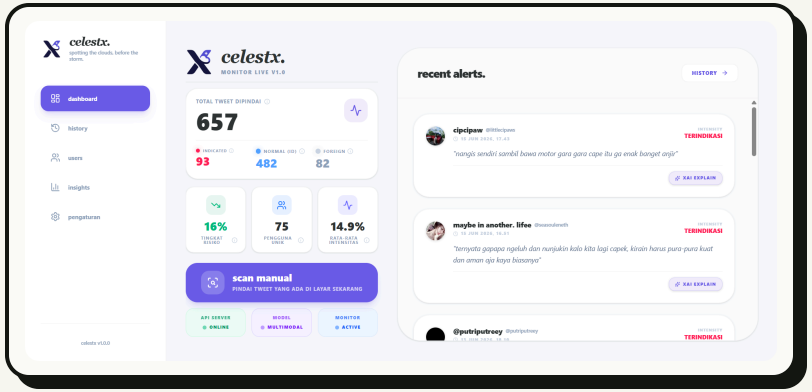



FIG. 3.2 · CLINICAL DASHBOARD

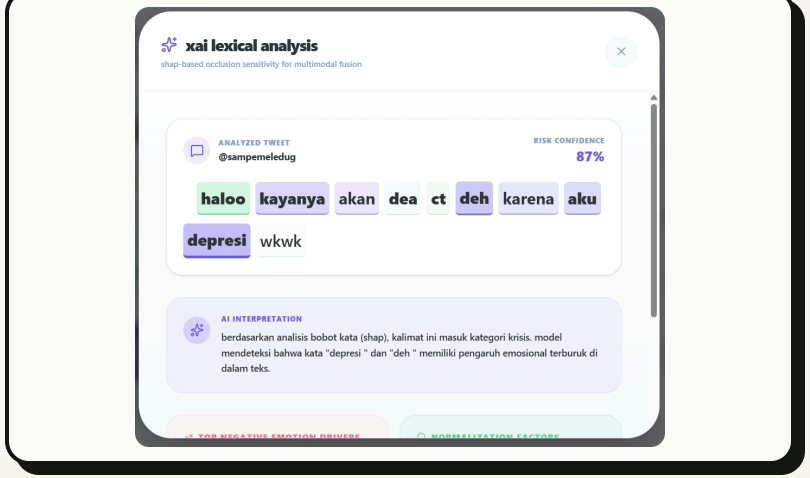



FIG. 3.3 · SHAP (XAI)

# BIG DATA MOVIE RECOMMENDER & LAKEHOUSE



DATA ENGINEER & BACKEND DEVELOPER

A real-time movie recommender: **Kafka** streams the dataset into a **MinIO** lakehouse, **Spark** builds the similarity model, a **Flask** web app serves similar-title picks, and a **Streamlit** dashboard monitors the stream live.

| 10,000+ | 7 | 5 MIN | TEAM (5) |
|---|---|---|---|
| MOVIE RECORDS STREAMED | PIPELINE STAGES | MODEL REBUILD LOOP | MEMBERS |

- **Streaming Ingestion:** A producer replays the dataset into the Kafka topic with duplicate prevention; Streamlit consumes the stream live with metrics and a chart.
- **Lakehouse & Model:** MinIO as object storage; Spark runs CountVectorizer + cosine similarity and exports the model as *.pkl* files.
- **Query & Web UI:** Flask serves similar-title search; the SQL layer uses *Trino + Hive Metastore + MariaDB*.

KAFKA · SPARK · MINIO · TRINO · STREAMLIT · FLASK · PYTHON · DOCKER

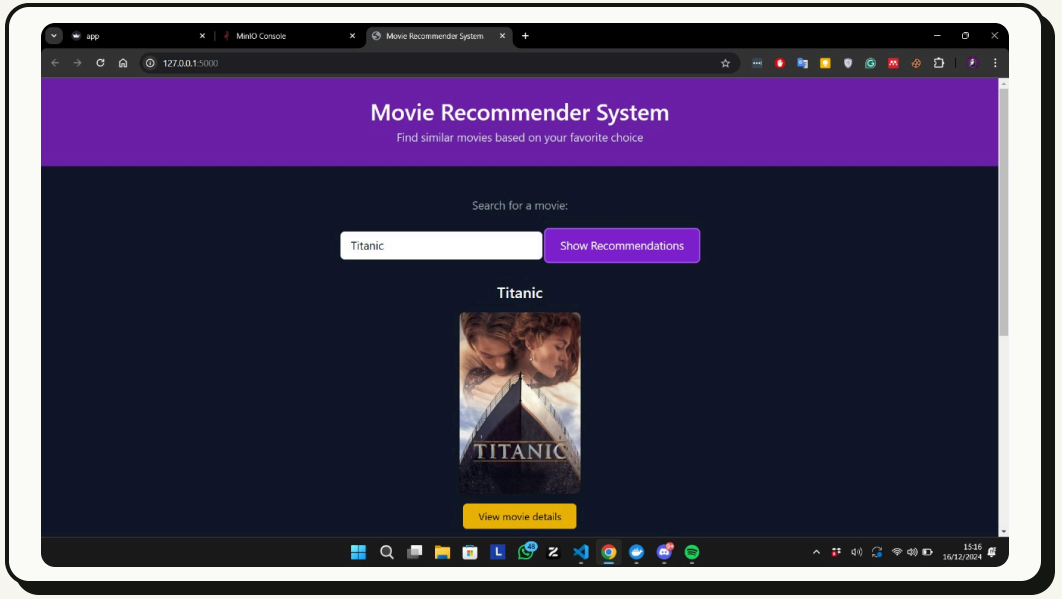



FIG. 4.1 · SIMILAR-TITLE RECOMMENDATIONS

# BIG DATA MOVIE RECOMMENDER & LAKEHOUSE



## STREAMING INGESTION & LAKEHOUSE

- **Producer:** replays *dataset.csv* into the Kafka topic, one record every half second, with duplicate skipping.
- **MinIO:** object storage (lakehouse); buffered rows are batched out as CSV and read back by Spark via *s3a://movies/*.csv*.
- **Spark:** tokenizes overviews, runs CountVectorizer, then computes a cosine similarity matrix exported as *.pkl* model files.

## QUERY, WEB UI & AUTOMATION

- **Flask:** serves the ten most similar movies with posters and ratings, plus IMDb links.
- **Trino + Hive Metastore + MariaDB:** the SQL layer over the lake, explored from DBeaver.
- **automation.py:** one script runs the producer, dashboard, Spark scheduler (every 5 minutes), and model monitor, orchestrated by Docker Compose.

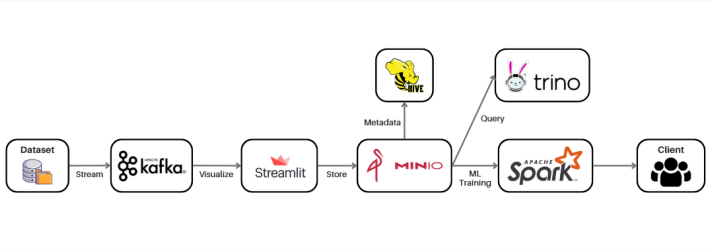



FIG. 5.1 · PIPELINE WORKFLOW

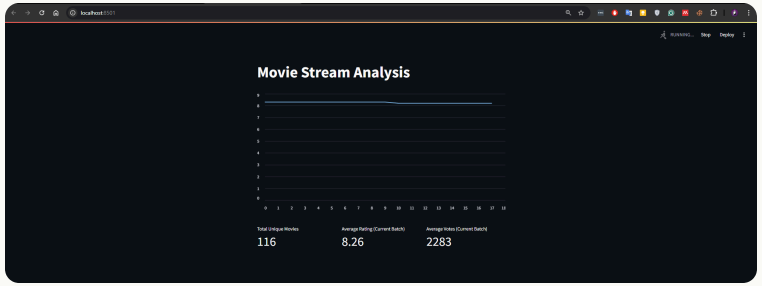



FIG. 5.2 · STREAMLIT LIVE MONITOR

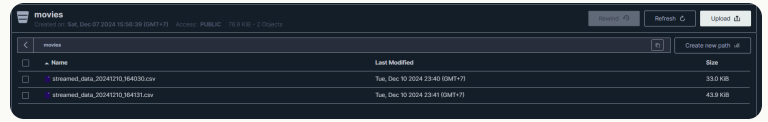



FIG. 5.3 · MINIO LAKEHOUSE

# BOOTHME



UI/UX & FULL-STACK DEVELOPER

A web-based AI photo booth: snap a selfie, and **Stable Diffusion 1.5 (revAnimated)** turns you into an illustrated character. Pick a frame, build a collage, and download the strip, with real-time **MediaPipe** face tracking.

| SD 1.5 | VUE 3 | TEAM (3) | 0 |
|---|---|---|---|
| REVANIMATED MODEL | REAL-TIME CLIENT | MEMBERS | INSTALLS · RUNS IN THE BROWSER |

- **No Install:** The whole experience runs in the browser; face tracking frames the shot before the shutter clicks.
- **Face-Aware Camera:** Real-time face detection via *Google MediaPipe*, so framing follows the subject.
- **Frame Studio:** Saved photos live in a gallery where you multi-select shots, pick a layout (e.g. 1x3), preview the collage, and download the strip.

STABLE DIFFUSION 1.5 · REVANIMATED · VUE 3 & VITE · MEDIAPIPE · NODE.JS (EXPRESS) · FIGMA · DOCKER

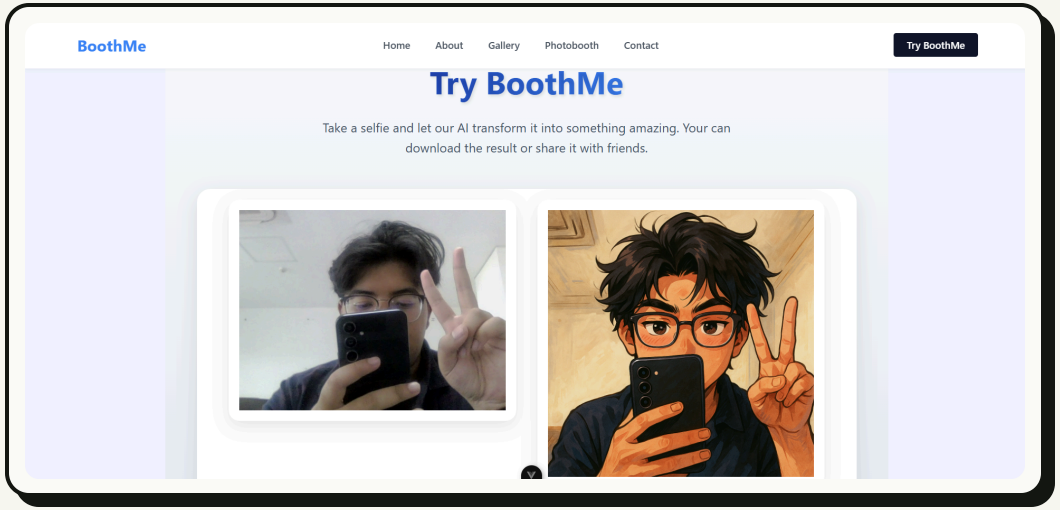



FIG. 6.1 · BEFORE AND AFTER

# BOOTHME



## BOOTH FLOW & FACE-AWARE CAMERA

- **Capture:** The Vue 3 client grabs a webcam frame and tracks the face with Google MediaPipe.
- **Encode & Generate:** The frame is packed as a base64 image, POSTed to Express, then proxied to Stable Diffusion WebUI on the GPU.
- **Return:** The illustrated result comes back to the client, ready for frames and collage.

## FRAME STUDIO & DEPLOYMENT

- **Frame Studio:** Saved photos in a gallery; multi-select, pick a layout (e.g. 1x3), preview a themed collage, download the strip.
- **Model revAnimated:** Stable Diffusion 1.5 tuned for semi-realistic digital illustration.
- **Docker Compose:** fe, backend, and sd webui services containerized to run consistently.

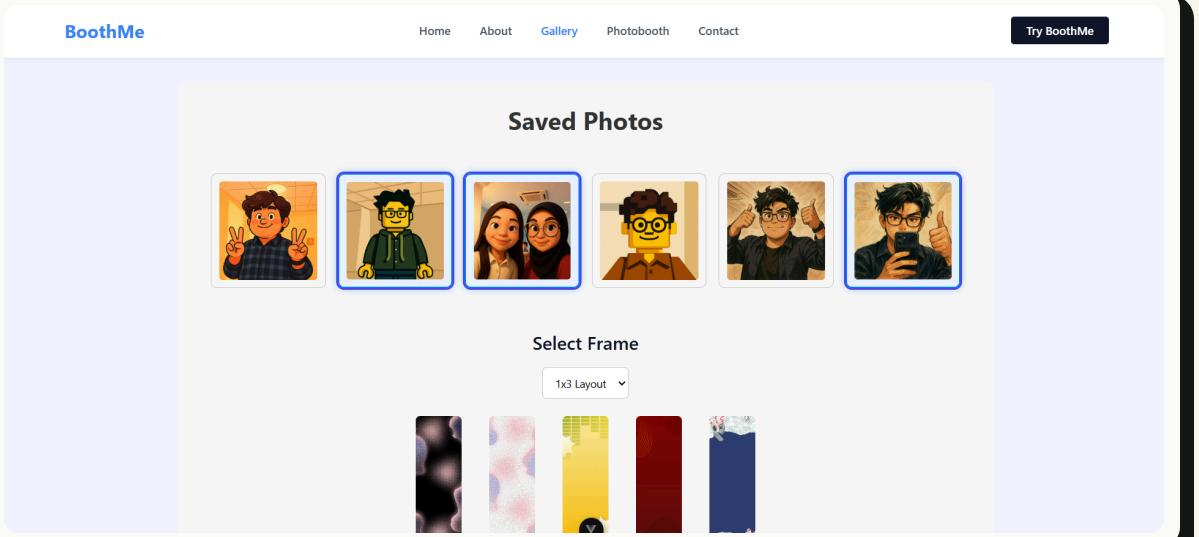



FIG. 7.1 · SAVED PHOTOS & FRAMES

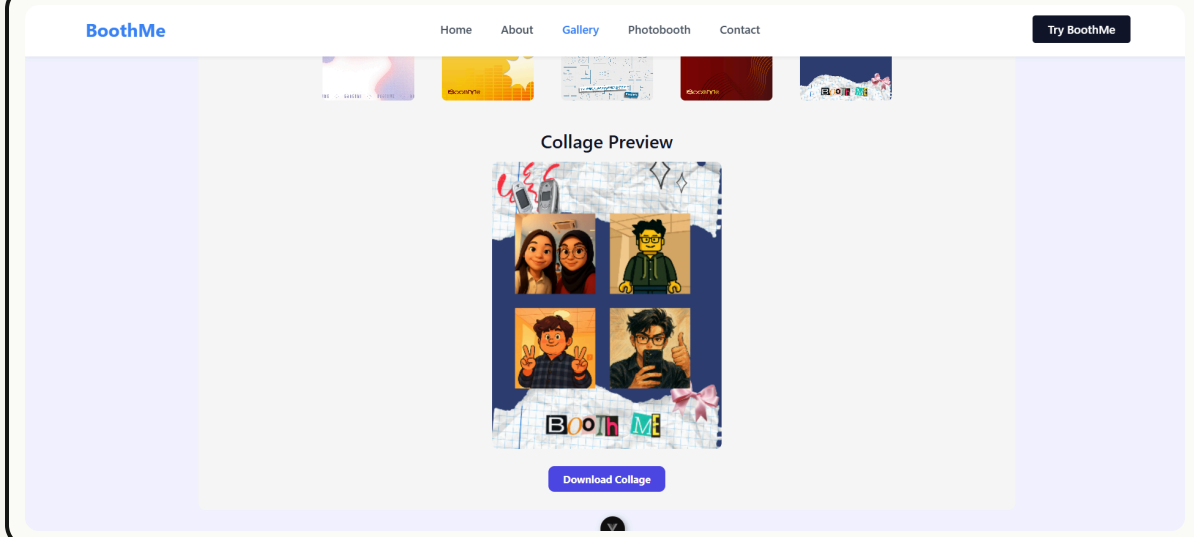



FIG. 7.2 · COLLAGE PREVIEW

# F1 RACE STRATEGY



DATA ANALYST · STATISTICAL MODELING · KAGGLE

F1 teams call an undercut "on" when the gap to the car ahead is smaller than the time a pit stop costs. Using **366,508 timed laps across 15 seasons**, I tested that rule and found it uninformative.

| 0.52 | 0.82 | 62%→6.6% | 57 |
|---|---|---|---|
| ROC-AUC · PIT-LOSS RULE ONLY | ROC-AUC · GAP + TYRE AGE | UNDERCUT CONVERSION BY GAP | TESTS · PYTHON & SQL AGREE |

- **The Problem:** The pit-wall shorthand says the pit loss decides the move, yet nobody runs it against 15 seasons of the archive.
- **What Decides It:** Once the rival stops too, both cars pay the pit loss and it cancels. Gap plus tyre offset reaches **ROC-AUC 0.82 out of sample**.
- **Where It Works:** Bahrain converts 49% of its undercuts, Monaco 7%, and a circuit's own pit loss does not order that table.

PYTHON · SQL (WINDOW FUNCTIONS) · SCIKIT-LEARN · PARQUET / SQLITE · PYTEST · KAGGLE

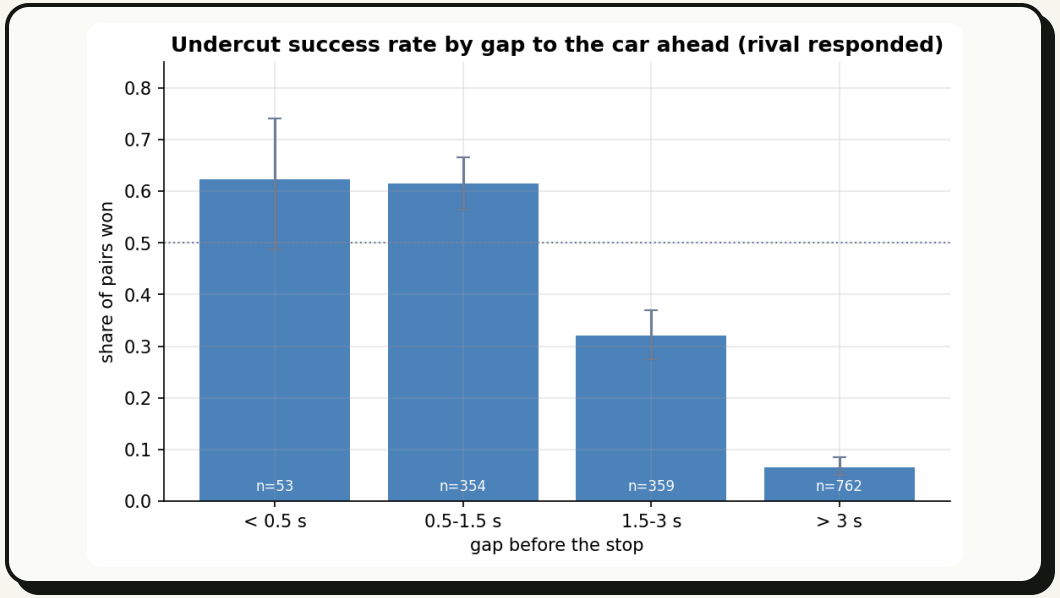



FIG. 8.1 · UNDERCUT SUCCESS BY GAP

# F1 RACE STRATEGY



## PIT LOSS & THE UNDERCUT

- **Estimator:** pit loss from lap times alone, *(in-lap + out-lap) − 2 × recent clean pace*; median 21.5 s across 36 circuits.
- **The cancellation:** both stops sit inside the window, so the pit loss drops out; what is left is *gain > gap*.
- **Validation:** the identity reproduces the classified order in 100% of the 1,528 responded pairs.

## PIPELINE & VALIDATION

- **Ingest & validate:** CC0 Ergast archive in, 24 automated quality checks; three real defects surfaced (810 durations published as lap times).
- **Two implementations:** 57 tests; Python and SQL window functions agree on all 36 circuits (0 disagreements over 366,508 laps).
- **Modelling:** pre-stop model ROC-AUC 0.847 in / 0.817 out of sample; next-lap stop prediction 0.796 ROC / 0.136 PR - a ranking model, not an alarm.

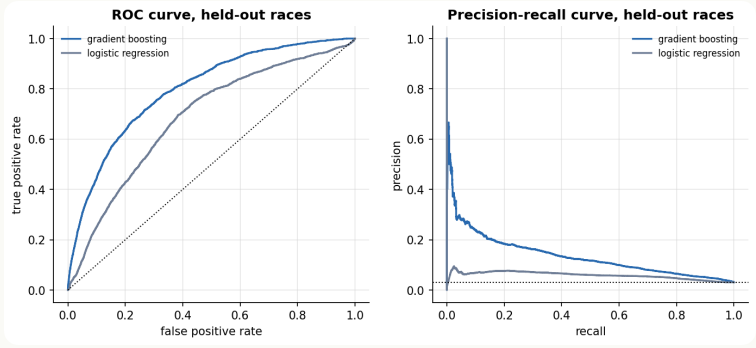



FIG. 9.1 · MODEL PERFORMANCE

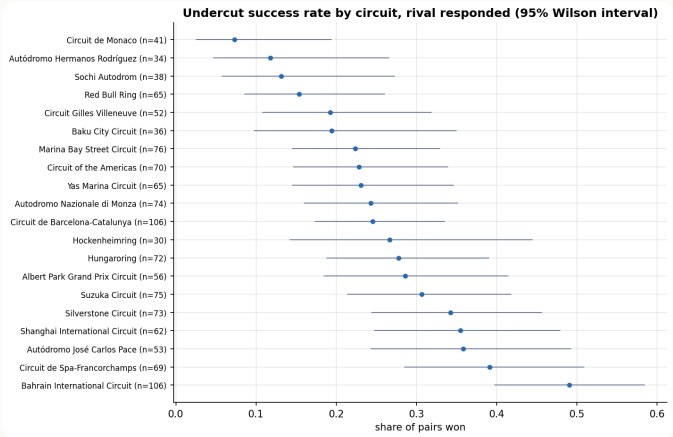



FIG. 9.2 · UNDERCUT BY CIRCUIT

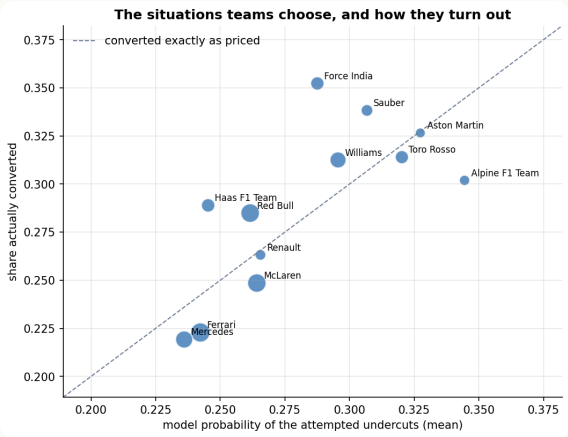



FIG. 9.3 · TEAM DECISION QUALITY

# RETAIL DEMAND FORECASTING



DATA SCIENTIST · END-TO-END BUSINESS ANALYSIS

A drugstore chain plans staffing, logistics, and stock per store per day. On the **million-row Rossmann dataset**, I measured how predictable daily sales really are and what a promotion is worth, forecasting with **PySpark Random Forest** and per-store target encoding.

**MAE $886**
DAILY SALES · UNSEEN JUN-JUL 2015

**$R^2$ 0.837**
58,611 TEST ROWS · RMSE $1,256

**+38.8%**
PROMO LIFT · $5,929 → $8,228

**844,392**
STORE OPEN-DAYS ANALYSED

- **The Problem:** Staffing and stock decisions need a daily baseline per store, plus an honest number for what promotions add on an ordinary day.
- **The Approach:** Trained on everything before June 2015 and scored on the two unseen months after, with target encoding fit on the training split only to keep leakage out.
- **The Result:** MAE $886 per store-day ($R^2$ 0.837); feature importance is led by the store's own baseline, then promotion and calendar.

PYSPARK · SPARK MLLIB · RANDOM FOREST · PANDAS · MATPLOTLIB

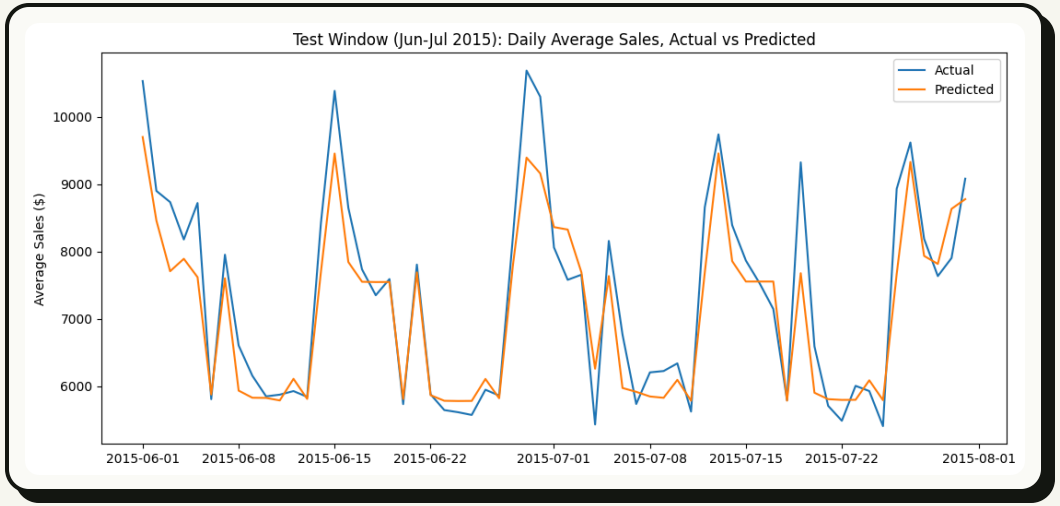



FIG. 10.1 · ACTUAL VS PREDICTED, UNSEEN WINDOW

# RSA-SECURED LOGIN FLOW



A 3-month internship at PT Badak NGL (Bontang, East Kalimantan): built an **RSA login flow prototype** for the company's internal systems e-Procurement, e-PNK and e-Contract, which share one login form and are used by the relevant departments. Tested with dummy data, then handed the source code to the IT team and demoed it for implementation.

## THE RISK

**CREDENTIALS OFTEN LIVE INSIDE INTERNAL DATABASES AS PLAINTEXT. IF THEY LEAK, EVERY ACCOUNT BEHIND THEM IS EXPOSED.**

## THE APPROACH

- Built **RSA asymmetric encryption** into the login flow: the public key encrypts credentials while they move through the login path, with database storage left to the later implementation.
- Keypair generation, login-flow wiring, and credential encryption end to end.

## DEVELOPMENT

- Built and tested with **dummy data**, then handed the source code to the IT team and presented and demoed it for implementation.
- Verified the credential stays encrypted from form input through to the point where the flow hands it off.

**3 MONTHS**
INTERNSHIP · DEC 2024 - FEB 2025

**RSA**
ASYMMETRIC CRYPTO

**3**
SYSTEMS · E-PROCUREMENT · E-PNK · E-CONTRACT

**SOURCE**
CODE HANDED TO THE IT TEAM

RSA KEYPAIR | LOGIN FLOW | CREDENTIAL ENCRYPTION | SOURCE CODE HANDOFF

# RSA-SECURED LOGIN FLOW



## THE CRYPTO FLOW

- **Login:** Credentials come from the form (message *m* = plaintext credentials).
- **Encrypt:** The RSA public key (e, n) encrypts: *c = m^e mod n*.
- **Scope:** The prototype stopped at the login flow; storing credentials in the database belongs to the later implementation.

## WHAT I BUILT

- **RSA Encryption:** RSA encryption for login credentials, including keypair generation.
- **Login Flow:** Wired encryption into the login flow, so credentials are encrypted before moving through the system.
- **Hand-off:** Source code handed to the IT team, who carried the flow into the internal systems.

```
login form
   |  m = plaintext credentials
   ▼
c = m^e mod n        # RSA public key (e, n) encrypts
   |
   ▼
hand-off             # source code handed to the IT team
```

# CLIENT DATA & ML DELIVERY



Project-based external engineer at **JokiProyek**: clients bring whatever they need built, from scraped and cleaned data to a RAG chatbot, and it goes back working. Remote contract, ongoing since August 2026.

## THE CONSTRAINT

**PROJECT WORK HAS NO FIXED STACK AND NO ROOM FOR A VAGUE ESTIMATE: SCOPE AND EFFORT ARE AGREED IN WRITING BEFORE THE FIRST LINE OF CODE.**

### SCOPE & BUILD

- **Scope of work:** every client request becomes a written scope and a delivery estimate before work starts.
- **Whatever the request is:** scraping and cleaning data, building the pipeline, or training the model, in **Python** (Pandas, Scikit-learn, PyTorch).

### DELIVERY & FIXES

- **AI when the job calls for it:** retrieval-augmented (RAG) chatbots and trained models alike.
- **Working tools and fixes:** FastAPI backends with Streamlit or React in front, plus debugging and bug fixing on existing client codebases.

**AUG 2026**
START · PRESENT

**SOW**
WRITTEN SCOPE PER PROJECT

**PYTHON**
DATA & AI WORK

**WEEKLY**
PROGRESS UPDATES

SCOPE OF WORK | DELIVERY ESTIMATES | PYTHON | RAG | FASTAPI | DEBUGGING

# HMIT ITS WEB PORTAL & DESIGN SYSTEM



The official web portal of HMIT ITS, designed as part of the association's **Website Development Team** (UI/UX division). Handled the design end to end: user research, interaction flows, and a component-based design system the front-end team built from.

## THE BRIEF

**ASSOCIATION NEWS, EVENTS, AND SERVICES WERE SCATTERED AND HARD TO NAVIGATE; THE REDESIGN RESTRUCTURED THE NAVIGATION SO ESSENTIAL INFORMATION IS REACHABLE IN A FEW CLICKS.**

### THE PROCESS

- Ran **user interviews** with students to map what they actually needed from the association's website.
- Restructured the **information architecture and user flows** so key sections are easy to find.

### DESIGN SYSTEM & HAND-OFF

- Built **design tokens** (color, typography, grid) and a reusable component library (buttons, cards) in Figma.
- Delivered **high-fidelity clickable prototypes**, then partnered with front-end engineers for pixel-perfect handoff into production components.

**2024**
WEBSITE DEVELOPMENT TEAM

**FIGMA**
DESIGN TOKENS & COMPONENTS

**HI-FI**
CLICKABLE HIGH-FIDELITY PROTOTYPES

**END-TO-END**
RESEARCH TO HAND-OFF

USER RESEARCH · UI/UX DESIGN · DESIGN SYSTEM · FIGMA · USER FLOWS · WIREFRAMING · PROTOTYPING

# OTHER WORKS



## FOOTBALL XG MODEL

MACHINE LEARNING · 2026

Kaggle notebook: an xG (expected goals) model over spatio-temporal shot coordinates, from distance and angle feature engineering to a live-match simulation sandbox.

PYSPARK · RANDOM FOREST · PANDAS · MATPLOTLIB

VIEW ON KAGGLE ↗

## FLIGHT DELAY PREDICTION

MACHINE LEARNING · 2026

Kaggle notebook: regression on US flight records to predict arrival delays in minutes.

PYSPARK · RANDOM FOREST · PANDAS · MATPLOTLIB

VIEW ON KAGGLE ↗

## MENTAL HEALTH INDO DATASET

DATASET · 2026

Public dataset release: 7.9k labeled Indonesian tweets for depressive-symptom detection, LLM-augmented and clinically validated (CC BY-NC-SA 4.0).

NLP · LLAMA-3.3-70B

VIEW ON KAGGLE ↗

# OTHER WORKS



## TRAFFIC SIGN CLASSIFIER

COMPUTER VISION · 2025

Course final project: an Android app that reads traffic signs from the camera feed and classifies them with a trained CNN.

ANDROID · JAVA · OPENCV · TENSORFLOW LITE

VIEW ON GITHUB ↗

## AR SHOOTING GAME

AR · 2025

Team project (2): an AR shooter for Android. Point the camera around your room, find the 3D targets, and blast them.

ANDROID · JAVA · ARCORE · SCENEFORM

VIEW ON GITHUB ↗

## LOCALINK

MOBILE · 2024

Team project (3): a location-based story sharing app. Built the main page, navigation, profile, and the story and location flows.

ANDROID · JAVA · FIREBASE · CAMERAX

VIEW ON GITHUB ↗

## WEB PROGRAMMING LAB MODULE

OPEN SOURCE · 2024

Contributed to the official web programming practicum module at ITS, including the responsive layout material in the HTML & CSS chapter.

HTML · CSS

VIEW ON GITHUB ↗

# HMIT ITS



The Information Technology Student Association (HMIT) at Institut Teknologi Sepuluh Nopember, Surabaya. Joined the **Media & Information department** as staff, then promoted to **Vice Head of Department**.

01

### DIGITAL PUBLICATION

Ran digital publication and visual communication campaigns across the department's channels, growing audience reach and engagement.

02

### LEADERSHIP

Promoted to Vice Head of Department: led and mentored junior designers, owning deadlines, creative direction, and asset QA.

03

### BRAND GUIDELINES

Authored organization-wide brand style guidelines, keeping print and online PR materials consistent.

**STAFF → VICE HEAD**
MEDIA & INFORMATION

**PUBLICATION CAMPAIGNS**
SOCIAL MEDIA & PR

**BRAND GUIDELINES**
UNIFIED VISUAL IDENTITY

**2024-2026**
FEB 2024 - FEB 2026

PUBLICATION · BRAND GUIDELINES · TEAM LEADERSHIP · VISUAL COMMUNICATION · SOCIAL MEDIA · MENTORING

# SELECTED DESIGNS



Poster covers, announcements, and weekly content designed for HMIT ITS publication channels, part of the Media & Information department's visual communication output.




FIG. 16.1 · CABINET COVER




FIG. 16.2 · INTERNSHIP ANNOUNCEMENT

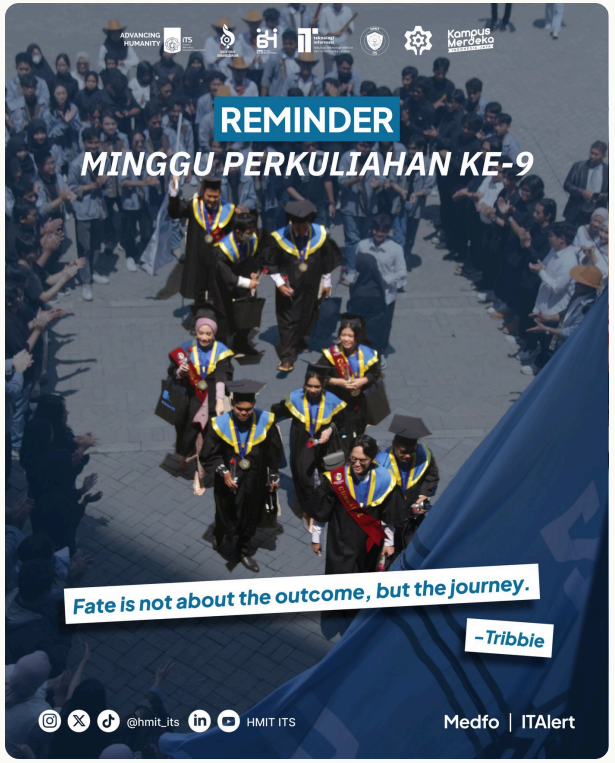



FIG. 16.3 · WEEKLY REMINDER

# TECHNICAL MATRIX



## DATA ANALYTICS & DATABASES

**Core tools:** Python, SQL (PostgreSQL, MySQL), Microsoft Excel (Pivot Tables, VLOOKUP, Power Query), Pandas, NumPy, Scikit-learn.

**Analytics:** Exploratory Data Analysis (EDA), statistical testing, data cleaning, data modeling, Matplotlib, Seaborn.

## MACHINE LEARNING & AI

**Frameworks:** PyTorch, Hugging Face Transformers, Stable Diffusion 1.5, Scikit-learn.

**Domains:** Natural Language Processing (IndoBERTweet), Computer Vision (Vision Transformer / ViT, MediaPipe), multimodal fusion, explainable AI (SHAP).

## BIG DATA & DATA ENGINEERING

**Streaming & storage:** Apache Kafka (real-time message streaming), Apache Spark, MinIO (S3 object storage lakehouse).

**Distributed query:** Trino distributed SQL engine, MariaDB, data partitioning, ETL pipeline design.

## FULL-STACK & SOFTWARE ENG.

**Backend:** FastAPI, Flask, Node.js (Express), RESTful APIs, OOP architecture, PHP, C++.

**Frontend:** Vue 3, Vite, React.js, Streamlit, HTML5, CSS3/Tailwind, JavaScript, cross-platform mobile (Flutter).

## DEVOPS & CLOUD TOOLS

**Infrastructure:** Docker, Docker Compose, Linux/Bash scripting, Git, GitHub version control, CI/CD basics.

**Testing & APIs:** Postman API testing, Chrome DevTools, Jupyter Notebooks.

## UI/UX DESIGN & LEADERSHIP

**Product design:** Figma high-fidelity prototyping, wireframing, user flow, brand guidelines, design tokens.

**Leadership:** ITS Teaching Assistant (4 semesters, 80+ students/semester), Vice Head, Media & Information (HMIT ITS).

## EDUCATION

S.Kom Information Technology · ITS Surabaya · 2022-2026 · GPA 3.62 / 4.00 · focus Data Science, AI-ML & Big Data

## TEACHING ASSISTANT

ITS · 4 semesters (Feb 2023 - Jun 2025) · 80+ students per semester · Big Data & Data Lakehouse, Web + Mobile Programming, Data Structures & OOP

## CERTIFICATIONS

IELTS Academic 7.0 (L 8.0 / R 8.0 / W 6.5 / S 6.0 · CEFR C1) · First author, ICAICTA 2026 (presented 20 Oct 2026)



OPEN TO WORK · JAKARTA / REMOTE

# LET'S TALK

Open for Data Scientist, Machine Learning Engineer, Data Analyst, or Software Engineer positions.

RADITHYARKAA@GMAIL.COM
LINKEDIN ↗
GITHUB ↗
KAGGLE ↗
WHATSAPP ↗