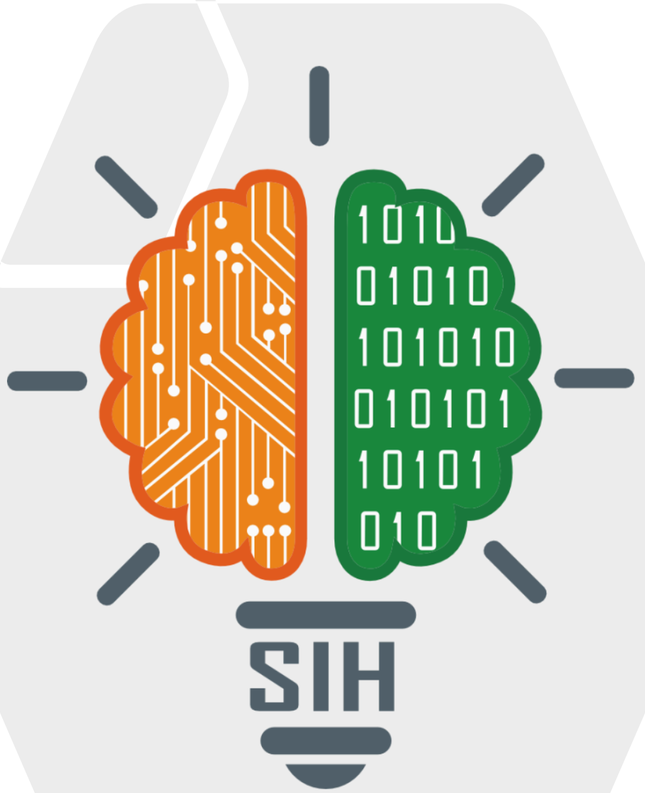
SMART INDIA HACKATHON 2025

SMART INDIA HACKATHON 2025



- **Problem Statement ID –** 25034
- **Problem Statement Title-** AI-Based Internship Recommendation Engine for PM Internship Scheme
- **Theme-** Education & Skill Development
- **PS Category-** Software
- **Team ID-**
- **Team Name -** SudoCode

SudoCode

Internship Recommendation Engine

SMART INDIA HACKATHON 2025

Proposed Solution:

- Lightweight AI-driven recommendation engine that suggests 3–5 best-fit internships.
- Considers skills, education, interests, location, and inclusivity needs.
- Optimized for mobile-first, low-literacy users with regional language support.

Innovation & Uniqueness:

- Multi-parameter scoring framework (20 signals → relevance, fairness, inclusivity).
- Hybrid engine: rule-based + ML-light, ensuring explainability + low compute.
- Bias-aware & diversity-aware recommendations (avoid repetition, ensure fairness).

SudoCode

TECHNICAL APPROACH

SMART INDIA HACKATHON 2025

Stack:

- **Language**: Python
- **API**: FastAPI + Uvicorn, Pydantic
- **DataBase**: PostgreSQL
- **Parsing**: PyMuPDF (for PDFs) + pytesseract (for scans/images)
- **Packages**: NumPy, pandas, dateparser, rapidfuzz, Haversine util
- **Auth/Security**: passlib[argon2] & TLS
- **Scheduler**: APScheduler (backend)
- **UI**: Streamlit (reads FastAPI)

AI / ML Models:

- **Document Type Classifier (for uploads)**:
  - Model: Logistic Regression (scikit-learn)
  - Inputs: TF-IDF over document text
  - Labels: {aadhaar, certificate, transcript, offer}
- **Recommendation "Nudge" (ranker)**:
  - Model: Logistic Regression (scikit-learn) — CPU-only, explainable
  - Training target: clicked/applied = 1, else 0 (class-weighted)

SudoCode

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2025

Feasibility:

- Lightweight, rule+ML hybrid → low compute + explainable
- Works in low bandwidth + low literacy environments
- Easily integrable into existing PM Internship portal

Challenges & Risks:

- Multilingual input/output handling
- Preventing urban/elite bias in recommendations
- Scaling to lakhs of users

Mitigation:

- Open-source translation + pictorial UI
- Bias audits + fairness constraints in scoring
- Cloud + modular design for scaling

SudoCode

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2025

Impact:

- First-gen learners and rural youth get clear, relevant choices
- Bridges skill–opportunity gap, reducing wasted applications
- Builds trust in PM Internship scheme

Benefits:

- Social: Equal access for rural/tribal youth, women, PWD candidates
- Economic: Better employability + higher placement outcomes
- Educational: Practical exposure → smoother industry entry
- Govt.: Higher efficiency of internship allocation

SudoCode

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2025

Python - *https://docs.python.org/3/*
FastAPI - *https://fastapi.tiangolo.com/*
uvicorn web server - *https://uvicorn.dev/*
Pydantic - *https://docs.pydantic.dev/latest/concepts/validators/*
PostgreSQL - *https://www.postgresql.org/docs/current/intro-whatis.html*
PyMuPDF - *https://pymupdf.readthedocs.io/en/latest/*
pytesseract - *https://github.com/madmaze/pytesseract*
NumPy - *https://numpy.org/doc/stable/user/index.html*
pandas - *https://pandas.pydata.org/docs/user_guide/index.html*
dateparser - *https://dateparser.readthedocs.io/en/latest/*
RapidFuzz - *https://rapidfuzz.github.io/RapidFuzz/*
argon2 - *https://argon2-cffi.readthedocs.io*
APScheduler - *https://apscheduler.readthedocs.io/*
streamlit - *https://docs.streamlit.io/*
Scikit-learn - *https://scikit-learn.org/stable/getting_started.html*