GabayPoz

Exploring the best-fit college programs for Pozorrubians

An end-to-end data project by Data Engineering Pilipinas in partnership with the SK Federation of Pozorrubio and Millennials PH Pangasinan

GabayPoz

Collaborators Onboarding

Be part of the generation that doesn't just talk about change, we measure it, we understand it, and we build it.

GabayPoz

Agenda

- About Data Engineering Pilipinas and The Puso Project
- Challenge, problem and project goal
- User story diagram, method and team structure
- Expectations and acceptance criteria per team
- Project timeline, meetings and learning support
- Code of conduct, AI usage guideline and tools and workflows
- Recognition and success metrics

Be part of the generation that doesn't just talk about change, we measure it, we understand it, and we build it.

Data Engineering Pilipinas

Data Engineering Pilipinas is a community for data engineers, data analysts, data scientists, developers, AI / ML engineers, and users of closed and open source data tools and methods / techniques in the Philippines. Data Engineering Pilipinas is a PyData group.

The Puso Project

Puso, heart, love.
Using data for good.
No matter how small the impact might seem.

Challenge Background

Many students rely on advice from peers, family, or social media, which may not reflect their own strengths, interests and particular circumstances. As a result, students may enter programs that do not align with who they are or where they want to be. In communities with limited financial resources, every educational decision carries significant weight.

GabayPoz

The Problem

There is currently no accessible, personalized system that helps students connect their skills, interests and constraints to a realistic academic pathways.

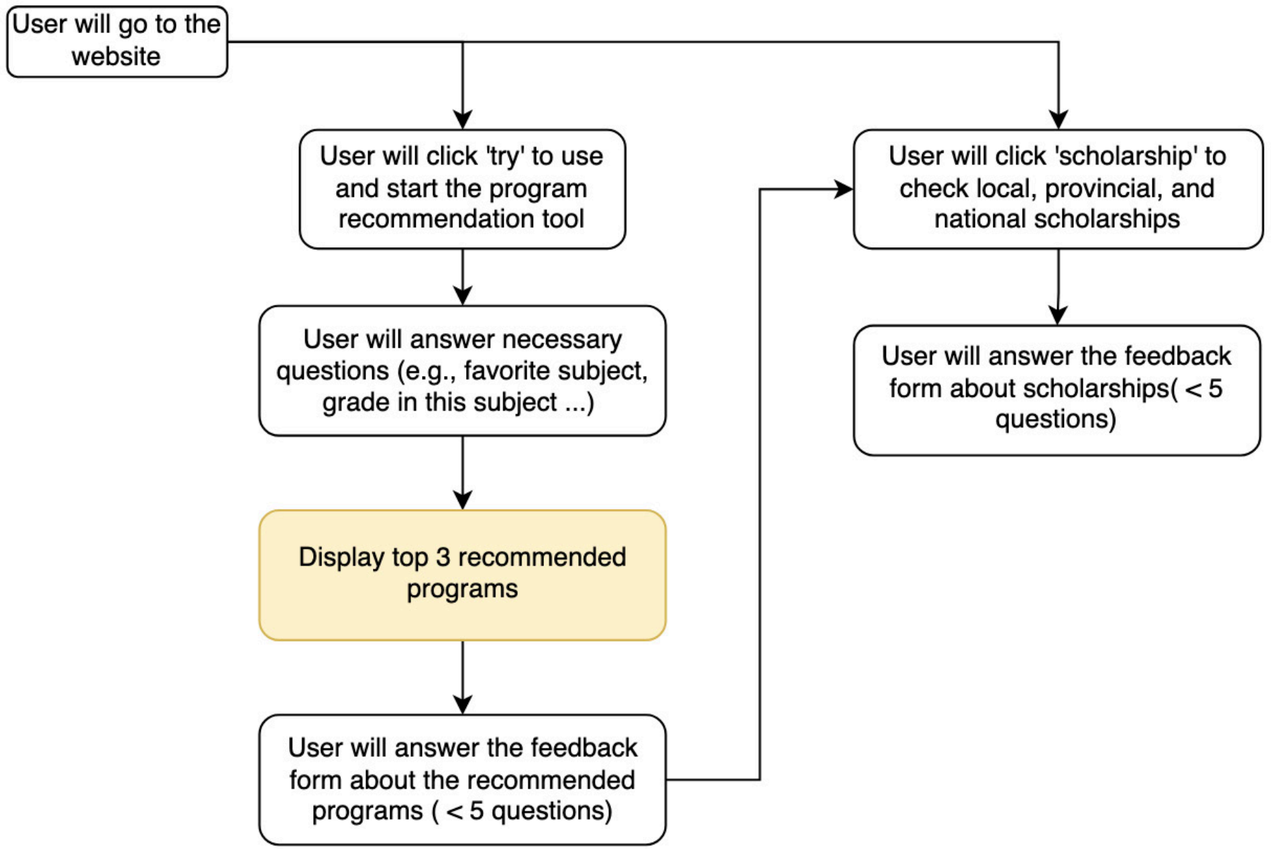
GabayPoz

Project Goal

The goal of the project is to develop a simple, data-driven digital-support tool that helps high school students, particularly Grade 11 and 12 identify college programs that best match their profile. Additionally, this will be an ecosystem for student decision-making as local, provincial and national scholarships will be included in the website as well.

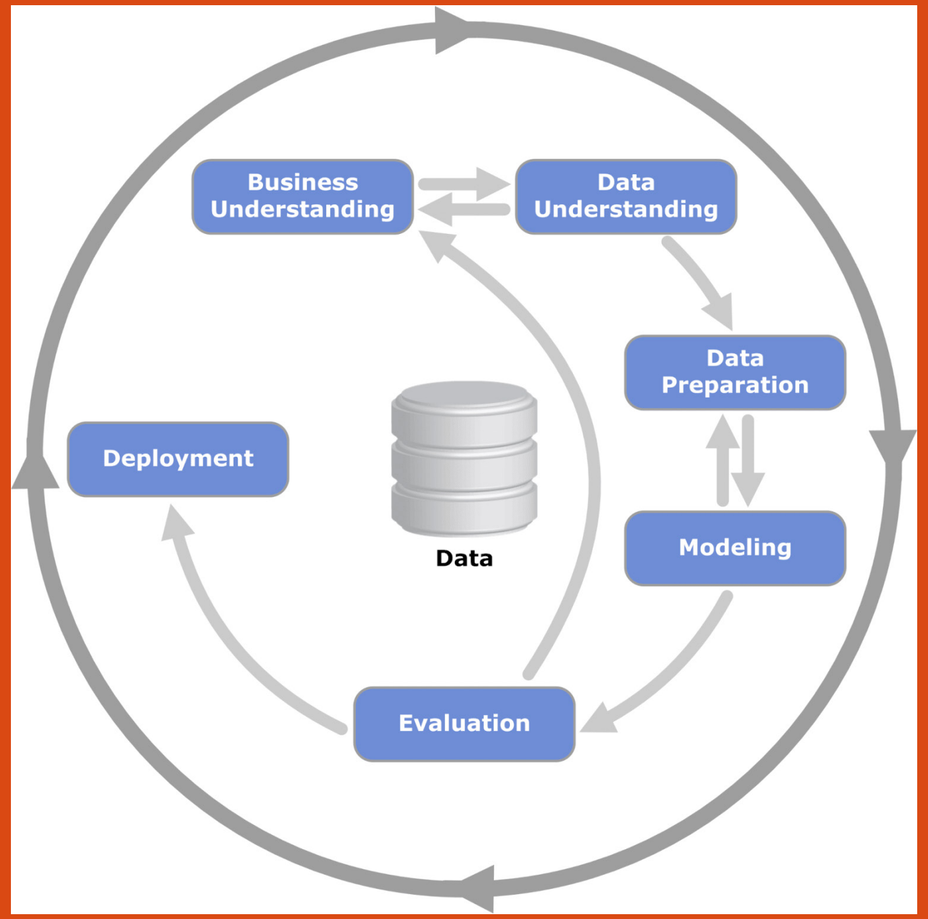
GabayPoz

High level user story diagram



GabayPoz

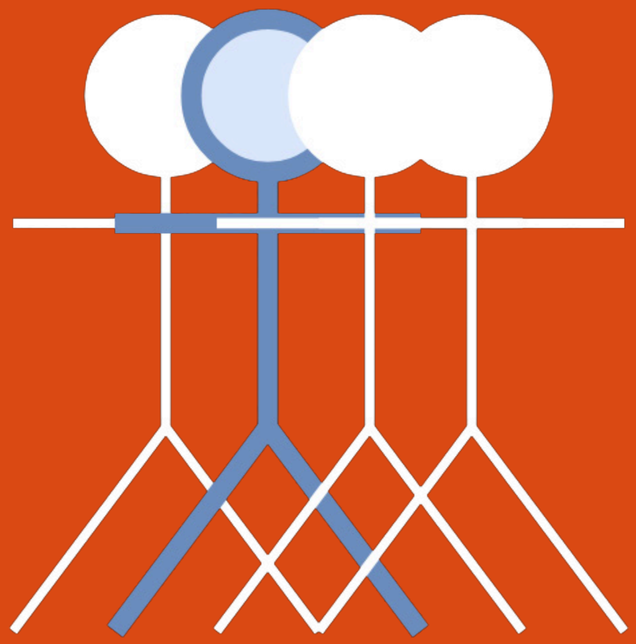
Project Method



Team 1 Knowledge
Team 2 Data Collection
Team 3 Data Engineering and EDA
Team 4 Model Development
Team 5 Web Development

GabayPoz

Team Structure



Each team should have one to two leads.

GabayPoz

Team 1

Knowledge

Collaborators will study existing literature, tools, and approaches related to career guidance systems, recommendation systems, and education support decision-making. The team should understand the factors that influence students' program choices, especially at the provincial level. They should also explore simple machine learning and rule-based approaches that can be applied to this project.

Team 1 Knowledge

Acceptance Criteria:

- Final project definition and scope
- A comprehensive review or research summary
- List of key variables that influence program recommendations
- Finalized framework for the student questionnaires
- Reference list (citations) of relevant literature, studies, tools, or datasets
- Post-survey questionnaire (after using the tool)

GabayPoz

Team 2

Data Collection

Collaborators will gather all relevant data needed to support the recommendation model. This includes employment and labor data, local industry information, list of nearby universities and their offered programs, market trend, agricultural data, literacy data, and any other related data. Data does not need to be perfectly structured.

GabayPoz

Team 2

Data Collection

Acceptance Criteria:

- A comprehensive documentation or data summary
- Data dictionary of each datasets
- Raw data in excel or csv format dump on Gdrive and git repo
- Program mapping
- List of scholarships available for the students including their requirements, deadlines, contacts and application process

GabayPoz

Team 3

Data Engg and EDA

Collaborators will clean, organize, and prepare the collected data to ensure consistency, accuracy and quality. The team will explore the dataset to identify patterns, distributions and relationships among variables. Data analysis will help validate whether the collected data is meaningful for a machine learning or just a rule-based system

GabayPoz

Team 3

Data Engg and EDA

Acceptance Criteria:

- Clean and structured dataset
- Data dictionary and schema documentation
- Exploratory data analysis
- Identification of important variables for modeling

GabayPoz

Team 4

Model Development

The team will design and implement the recommendation logic that recommends three college programs based on student responses and collected data. Depending on the data collected, the first version may use rule-based logic or scoring methods, while later versions may integrate simple machine learning algorithms such as clustering or similarity-based recommendations.

GabayPoz

Team 4

Model Development

Acceptance Criteria:

- Defined recommendation logic or algorithm which includes at least one-page document containing explanations and at least one python file implementing the final algorithm.
- Model documentation and testing results
- Final list of possible output (programs e.g., BSCE)
- Working model or prototype

GabayPoz

Team 5

Web Development

Collaborators will build the user interface that allows students to interact with the system. This includes creating the questionnaire interface, connecting it to the recommendation engine, and presenting the results in a clear and understandable format.

GabayPoz

Team 5

Web Development

Acceptance Criteria:

- Basic ui/ux design for ease of use
- Functional web interface for student input
- Integrated backend connection to recommendation logic
- Result page for the top 3 recommended programs
- Feedback questionnaires page
- Scholarship page as an additional feature

GabayPoz

Project Timeline

	Start Date	End Date
Team 1	March 29(Monday)	April 12 (Sunday)
Team 2	April 13 (Monday)	April 26 (Sunday)
Team 3	April 20 (Monday)	May 3 (Sunday)
Team 4	May 4 (Monday)	May 10 (Sunday)
Team 5	April 13 (Monday)	May 17 (Sunday)

Go-live: May 13, 2026

Length of the Project: 7-8 weeks

GabayPoz

Post-development Timeline

	Start Date
Go-live	May 13 (Wednesday)
Official Presentation to the Stakeholders	May 22 (Saturday)
Official Presentation to the public via discord, Facebook live and zoom	May 23 (Sunday)
Datamasters presentation team 1 and 2	May 30 (Saturday)
Datamasters presentation team 3 and 4	June 13 (Saturday)
Datamasters presentation team 5	June 27 (Saturday)

GabayPoz

Meeting Structure & Expectations

30-min weekly sync session
Standups session per team if needed
15-min handshake between team

Learning and Support

Asking questions is part of contributing.
No question is too basic.
Progress > Perfection.

GabayPoz

Code of Conduct

Respect and Inclusivity

No discrimination, harassment, or dismissive behavior.

Everyone is learning at different speeds.

Communication Guidelines

Be clear and respectful.

Give constructive feedback.

Avoid ghosting after taking a task

Responsibility & Accountability

If you take a task → give updates

If you can’t continue → inform early

GabayPoz

AI Usage Guideline

AI as a Co-Pilot not an Autopilot

- AI helps in coding, debugging, learning new libraries, improves documentation
- AI should NOT replace critical thinking

Prompt Transparency Rule

If AI is used for:

- Writing code, generating analysis, documentation

Then contributors must include:

- Prompt(s) used
- AI tool used (e.g., ChatGPT, Claude, Copilot)
- What was modified after AI output

GabayPoz

Discouraged Practices

- Submitting AI-generated work **without review**
- Using AI to **fabricate** results or analysis
- Copying outputs **without understanding**

Learning First Policy

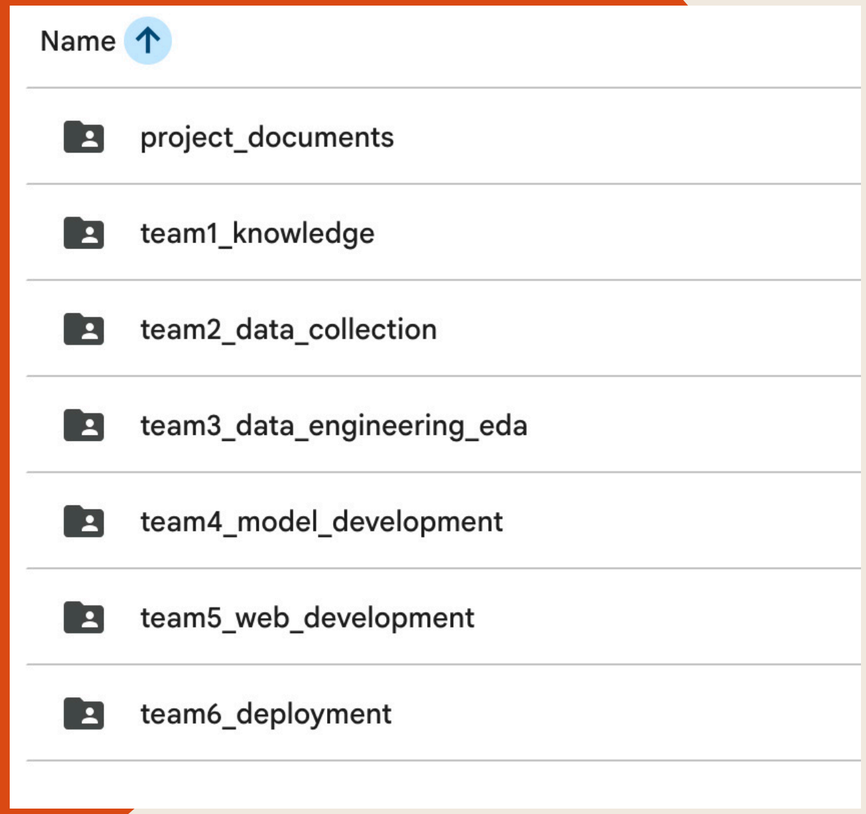
- Use AI to learn faster. It's more important to focus on your learning journey rather than sacrificing it to give quick results.

GabayPoz

Tools and Workflows

- Communication: Slack
- Code: GitHub
- Docs: Google Drive
- Meetings: Zoom / Google Meet

GabayPoz



File naming conventions

- use snake_case (lowercase and underscore)
- be descriptive and concise

Suggested Format:

<team>_<task>_<description>_v<version>.<ext>

Example:

team3_data_cleaning_work_data_v1.py

team3_eda_student_dropout_v2.ipynb

team4_model_recommendation_baseline_v1.pkl

team2_structured_data_agriculture_2024.csv

GabayPoz

Contribution and Recognition

How Contributions Are Tracked:

- Task completion
- Participation in discussions and meetings
- GitHub commits

Recognition Types

- Certificate
- Face will be included in the 'builder' page
- Name will be included in DEP-TPP volunteers

Portfolio Value

- Add the project to your resume
- Talk about it in interviews
- Showcase work on GitHub

GabayPoz

Success Metrics

Student Engagement

Stakeholder Feedback

Recommendation Performance

System Performance

GabayPoz

Next Steps

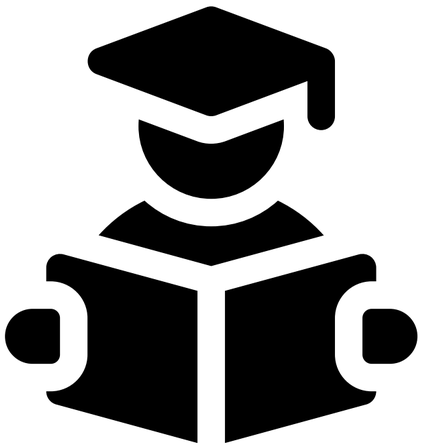
Join slack channel:

Access google drive:

GabayPoz

Picture taking

Be part of the generation that doesn't just talk about change, we measure it, we understand it, and we build it.



GabayPoz

GabayPoz