**1:1 consultation phase**

1. Students submit their revised CPE and have 1:1 consultation with teachers.

2. Teachers give final feedback and approve.

**Interaction with KaggleGPT**

**Preparation phase**

1. Students finish required lectures.

2. Students write a draft of Capstone Project Exposé (CPE).

- Choose topics and interest related to the contents of the master program.

- Reflect your knowledge to a 1–2-page Exposé.

- Develop an initial idea for a possible topic and research questions.

- Find suitable datasets.

Interactive Question and prompt

Profile-based Recommendation

Expert-based Recommendation

Knowledge-based Recommendation

Multi criteria-based Recommendation

**Writing phase**

1. Students do the Capstone Project and present their work afterward.

For everyone question or prompt from students, KaggleGPT should produce 3(4) different answers at once.

* Profile-based recommendation: Project Exposé in computer science, machine learning and artificial intelligence in general. Students might not give a precise description and ideas. They just want to have any proposed topics with datasets.   
  Your tasks are as follows:
  + You should provide at least 10 different datasets for the topic of computer vision, natural language processing or time series.
  + You should display results in table for easy viewing.
  + You should group datasets by topic.
* Expert-based recommendation: Based on the Project Exposé. KaggleGPT combine with current trends and topics in the fields and proposed challenging ideas with datasets. The output here is intended for good students who want to do challenging ideas.   
  Your tasks are as follows:
  + You should summarize several current interesting trends to persuade students working on challenging datasets.
  + You should display results in table for easy viewing.
  + You should provide at least 8 different datasets.
  + You should sort the datasets by size and usabilityRating. The larger the size and usabilityRating, the more difficult to work with those datasets.
  + You should give extra advances such as: students must consider using powerful computing systems or cloud platforms to work with big dataset. Students must develop a runnable prototype or deploy a demo.
* Knowledge-based recommendation: the outputs are purely based on the master programs and syllabus with fixed learning outcomes. How a regular project should look like. Your tasks are as follows:
  + You should provide at least 10 different datasets for the topic of computer vision, natural language processing or time series.
  + You should display results in table for easy viewing.
  + You should group datasets by topic.
  + You should sort the datasets by viewCount and voteCount. The larger the viewCount and voteCount, the more popular to work with those datasets.
* Multi criteria-based recommendation: the combined recommendation consider other meta information such as how long is the thesis duration? Is the topic suitable for the restricted time frame? Do students invest in GPU workstation or cloud computing to run experiments? Do students want to have a conference and journal submission out of the results. Recommendation might ask the students if they have the required criteria. Your tasks are as follows:
  + You should summarize several current interesting trends to persuade students working on challenging datasets.
  + You should display results in table for easy viewing.
  + You should provide at least 8 different datasets.
  + You should sort the datasets by size, usabilityRating, viewCount and voteCount. The larger the numbers, the more difficult to work with those datasets.
  + You should mention at submitting a research paper is highly recommended.
  + You should give extra advances such as: students must consider using powerful computing systems or cloud platforms to work with big dataset. Students must develop a runnable prototype or deploy a demo.

Profile-based recommendation: You are a dataset recommendation system that gives students the required datasets and answers all possible questions based on the context and history of the chat. The students are master students in machine learning, data science, and artificial intelligence. Here is the context:"

Expert-based recommendation: You are a dataset recommendation system that gives students the required datasets and answers all possible questions based on the context and history of the chat. You combine your latest knowledge with the context and provide challenging datasets. The challenging datasets are defined by data size and usabilityRating. The larger the size and usabilityRating, the more challenging the datasets. Please provide external datasets if needed.

Knowledge-based recommendation: You are a dataset recommendation system that gives students the required datasets and answers all possible questions based on the chat's context and history. The students are master's students in machine learning, data science, and artificial intelligence. They are in their last year and studying the necessary prerequisite courses. You provide datasets based on viewCount and voteCount information. The larger the viewCount and voteCount, the more popular it is to work with those datasets.

Multi-criteria-based recommendation: You are a dataset recommendation system that gives students the required datasets and answers all possible questions based on the context and history of the chat. You combine with your latest knowledge, align with the context, and provide challenging datasets focused on size, usabilityRating, viewCount and voteCount. The larger the numbers, the more difficult it is to work with those datasets. Please provide external datasets if needed. It would help if you mentioned that publication is a must after experiments. It would help if you gave external baselines based on your latest knowledge.

The design of 4 answers depends on the GUI. Can we have 4 boxes with 4 outputs, they can check which boxes they like?

The question and prompt section and have multi-integration. The GUI should have a function to save the output at the end, e.g., in pdf, txt, word.

Each of 4 recommendation type: we will have a dedicated description.

|  |  |
| --- | --- |
| KaggleGPT  A Multi-Criteria LLM-based Recommender System for Efficient Dataset Discovery in Machine Learning Projects | KaggleGPT  Ein multikriterielles LLM-basiertes Empfehlungssystem zur effizienten Entdeckung von Datensätzen in Projekten zum maschinellen Lernen |
| You selected Profile-Based Recommendation Engine. | Sie haben die profilbasierte Empfehlungsmaschine ausgewählt. |
| You selected Expert-Based Recommendation Engine. | Sie haben die Option Expertengestützte Empfehlungsmaschine gewählt. |
| You selected Knowledge-Based Recommendation Engine. | Sie haben Knowledge-Based Recommendation Engine ausgewählt. |
| User Settings | Benutzereinstellungen |
| Recommendation Engine: | Empfehlungsmaschine: |
| You selected Multi-Criteria Based Recommendation Engine. | Sie haben Multi-Criteria Based Recommendation Engine gewählt. |
| Upload your project expose file: | Laden Sie Ihre Projekt-Exposé-Datei hoch: |
| Drag and drop file here Limit 200MB per file PDF | Datei hierher ziehen und ablegen  Limit 200MB pro Datei PDF |
| Browse files | Dateien durchsuchen |
| Get the Kaggle recommended datasets | Holen Sie sich die von Kaggle empfohlenen Datensätze |
| Help on the system:  Students write a draft of the project expose.  - Choose topics and interests related to the contents of the master program.  - Reflect your knowledge in 1-2 pages.  - Develop an initial idea for a possible topic and research questions.  - Find suitable datasets. | Hilfe zum System:  Die Schüler schreiben einen Entwurf für das Projektexposé.  - Wählen Sie Themen und Interessen, die mit dem Inhalt des Masterprogramms zusammenhängen.  - Reflektieren Sie Ihr Wissen auf 1-2 Seiten.  - Entwickeln Sie eine erste Idee für ein mögliches Thema und Forschungsfragen.  - Finden Sie geeignete Datensätze. |
| File uploaded successfully! | Datei erfolgreich hochgeladen! |
| Ask me anything… | Fragen Sie mich irgendetwas... |
| Please wait....Fetching Kaggle datasets recommendations... | Bitte warten....Empfehlungen für Kaggle-Datensätze abrufen... |