

## Visvesvaraya National Institute of Technology (VNIT), Nagpur

# Personalized EdTech Mentor (An RAG Based Project)

## **Project Report**

Submitted by:

Chittoor Vamsi (MT23AAI069) Fahad Hassan (MT23AAI035) Md Arshad Hussain (MT23AAI015)

Submitted to:

Dr. Saugata Sinha

Department of Electronics and Communication Engineering, VNIT

Nagpur

#### **Problem Statement**

Students these days juggle slides, PDFs, articles and more—and it's tough to sort out what really matters. Most e-learning sites push everyone through the exact same modules and quizzes, no matter their background or pace. The result? Frustration, wasted time, and half-remembered lessons. What we need is a smarter helper: one that reads whatever you upload, pulls out the key ideas, quizzes you on your weak spots, answers your follow-ups in context, and tracks your progress so you can see exactly where you stand. This project builds that AI-powered study buddy—responsive to each student's unique pace and learning style.

#### **Abstract**

We've built the Personalized Education Mentor, a web app that turns your own PDFs—be they lecture notes, research papers or chapters—into tailored study sessions. After you upload a file, a language model breaks it down into topics you can explore. A vector database makes finding the right paragraph nearly instant, and a quiz engine generates questions targeting your trouble spots. As you work through the material, the app notes which concepts you've mastered and which need more attention, then suggests what to tackle next. Developed with Python, Streamlit and LangChain, and powered by GPT-4, HuggingFace embeddings and Chroma DB, our prototype shows how modern NLP tools can make independent learning both efficient and engaging.

## Introduction

Personalized learning isn't a luxury—it's essential. Every student brings their own mix of prior knowledge, interests and stumbling blocks, yet most online courses ignore that fact. Thanks to advances in NLP and machine learning, we can now flip the script: upload any PDF, and an LLM will parse it into bite-sized concepts. A vector store then lets you query specific sections on demand or dive into quizzes tailored to what you haven't yet grasped. If you hit a snag, the system pulls the answer straight from your material, in context. Over time, it quietly learns where you shine and where you need a bit more practice—then guides you toward the next topic. In the chapters that follow, we'll cover related work, walk through our system's architecture and implementation, and share the results from our initial tests.

#### **Tools Used**

- Python 3.x
- Streamlit
- LangChain
- OpenAI GPT-4
- HuggingFace Embeddings and Chroma DB (Vector Store)

## **Procedure**

**1. Data Collection :** To kick things off, students upload their own PDFs—lecture notes, textbook excerpts or journal articles—through the Streamlit UI.

Each file automatically lands in the data/ folder so we always have the original to fall back on.

**2. Text Segmentation :** A recursive splitter slices each document into neat, self-contained fragments (we avoid cutting sentences in half).

**The result:** bite-sized passages that remain meaningful on their own.

- **3. Embedding Generation :** Every fragment is converted into a semantic vector via HuggingFace's sentence-transformers. These vectors capture the "aboutness" of each passage and power our later searches.
- **4. Vector Store Construction :** We load all embeddings into Chroma DB, which serves as our fast similarity index. In practice, a query against Chroma returns top matches in milliseconds.
- **5. Topic Discovery :** GPT-4 combs through the indexed fragments to pull out major topics and subtopics.

That topic hierarchy is saved in JSON format, giving users a clear menu of study options.

- **6. Dialog-Driven Q&A:** When a learner picks a topic, the system fetches the most relevant text pieces from Chroma. GPT-4 then weaves those snippets into concise, context-aware answers—keeping prior questions in mind.
- 7. Quiz Crafting & Feedback: For each topic, the app autogenerates multiple-choice questions using only the educational text (no headers, no captions).

Students answer right in the UI and instantly see which items they got right or wrong.

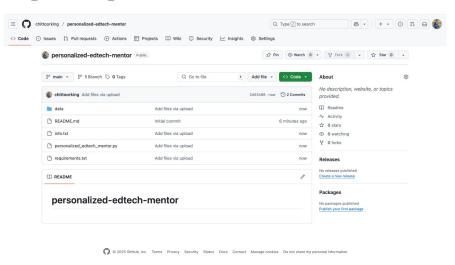
**8. Performance Tracking & Recommendations :** Every quiz result updates a per-topic mastery profile.

Based on that profile, the system flags concepts needing review and suggests the best next steps—so each learner's path stays personalized.

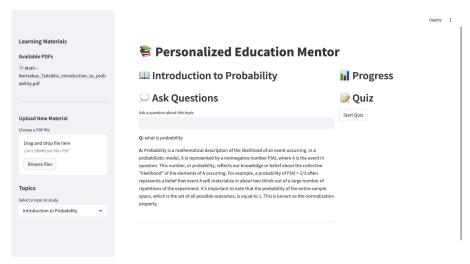
#### **Results**

The Personalized Education Mentor was successfully implemented as a web application. Users can upload their own PDFs, extract topics, ask questions, and take quizzes. The system provides immediate feedback and tracks progress for each topic. The following images illustrate the application's interface and workflow:

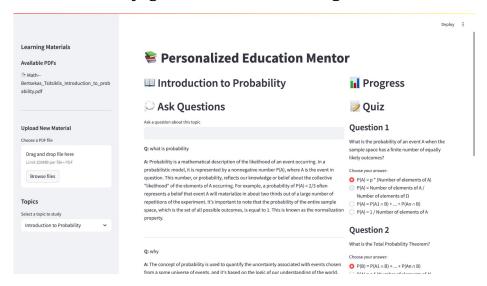
#### Repo setup of the code



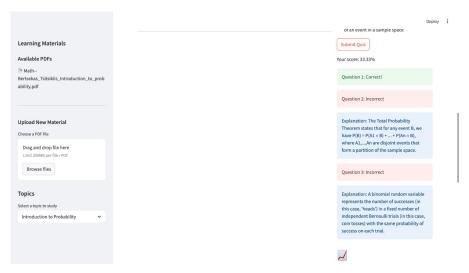
**1. Chat Interface:** User interacts with the bot to ask questions about the selected topic.



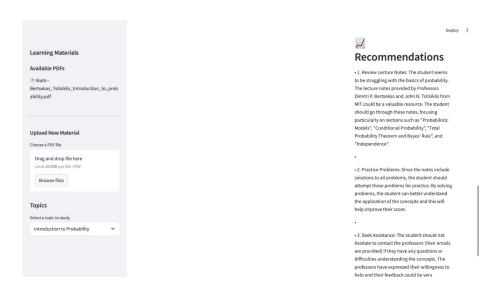
**2. Dynamic Quiz Generation:** User starts quiz with questions automatically generated on the right side of the screen.



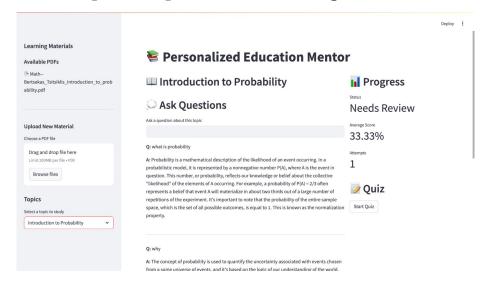
**3. Score & Explanations:** User receives quiz score and detailed explanations for incorrect answers.



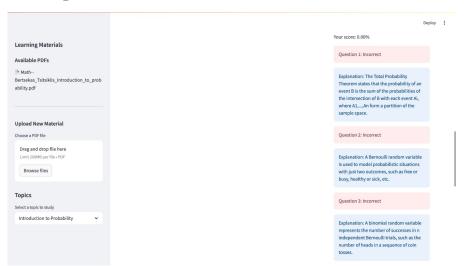
**4. Smart Recommendations:** System provides personalized recommendations based on all previous attempts for the topic.



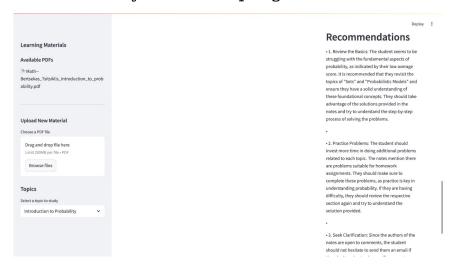
**5. Progress Tracking:** Returning users see a progress card showing average score and total attempts count. Users can reattempt to improve their average.



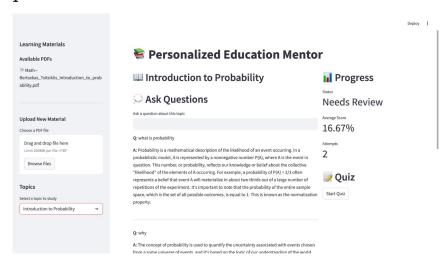
**6. Score Impact:** Incorrect answers lower the average score for the topic.



**7. Adaptive Recommendations:** Recommendations update automatically based on progress card data and attempt history.

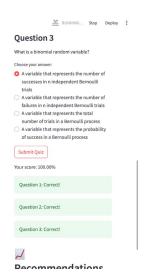


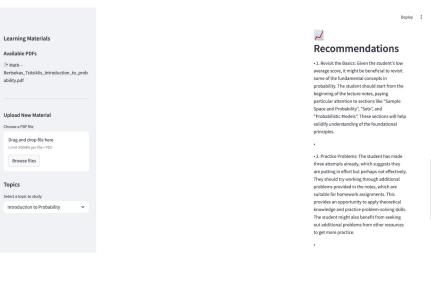
**8. Score Decline:** Average score decreases after poor performance.



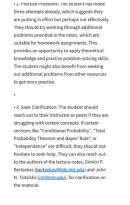
**9. Perfect Score Handling:** All correct answers generate no explanations but still update recommendations based on latest progress.



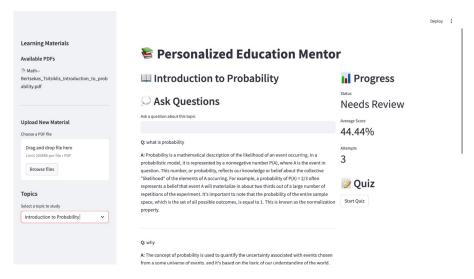




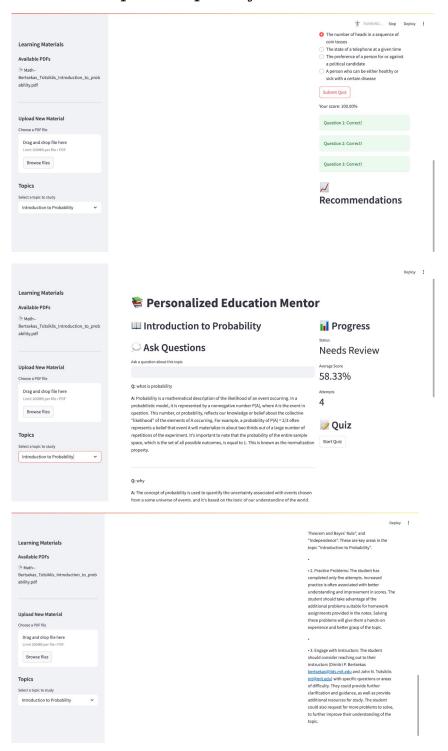




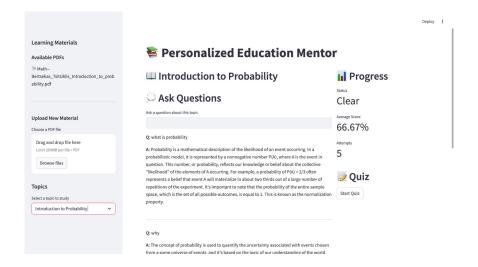
## **10. Score Improvement:** Average score increases with better performance.



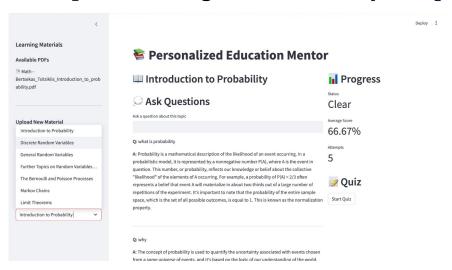
**11. Continuous Learning:** Users can reattempt quizzes until they clear the topic completely.



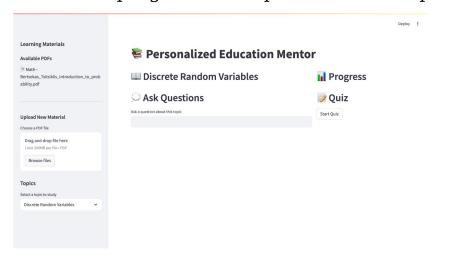
**12. Topic Mastery:** After 5 attempts, the score improved and status changed to "clear"



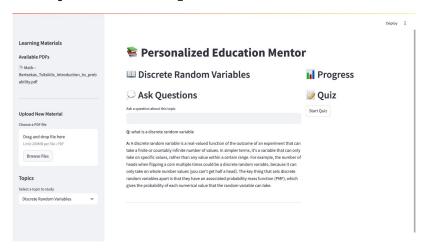
13. Topic Switching: Users can easily change topics at any time.



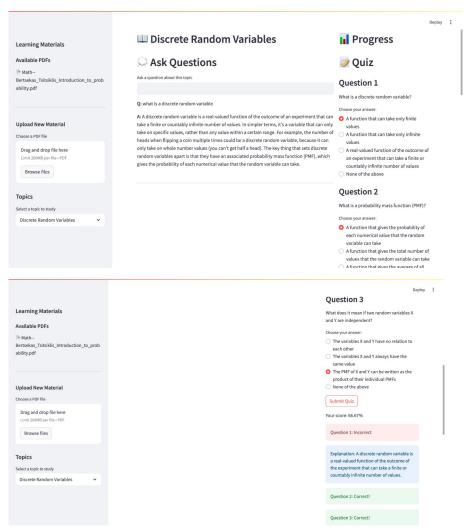
**14. Fresh Start:** New topic selection provides fresh chat history and a new progress card specific to that topic.

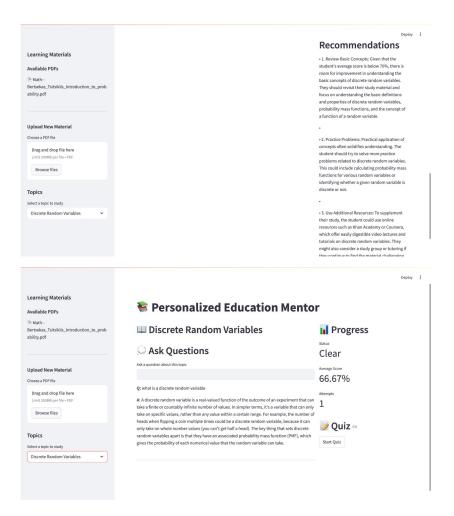


**15. Topic-Specific Help:** Users can ask questions about their newly selected topic.

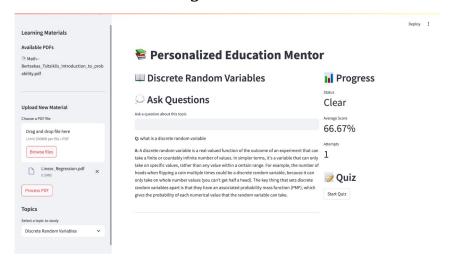


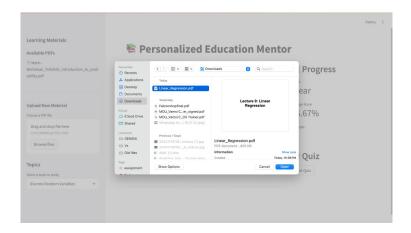
**16. Consistent Process:** Same quiz workflow applies - attempt until cleared with explanations and recommendations provided.



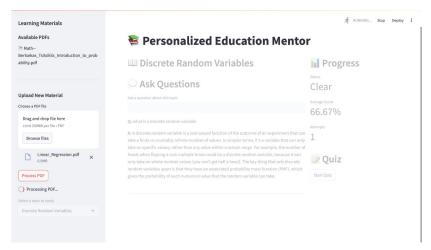


**17. PDF Upload:** Users can upload new PDFs by clicking "Browse Files" and selecting documents.

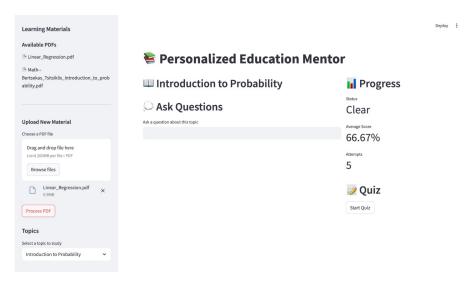




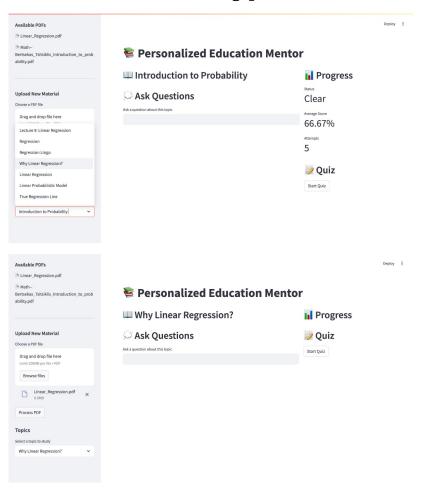
**18. Auto-Processing:** New PDFs are embedded into ChromaDB and topics are automatically updated after clicking "Process PDF."



**19. Database Integration:** Newly uploaded PDFs appear in the available documents list within ChromaDB.



**20. Seamless Integration:** Users can select new topics and follow the same learning process as before.



## **System Architecture**

#### **Overview**

This system helps learners by letting them upload study materials and ask questions about them. It also creates quizzes to test their understanding. The system is designed to work entirely from the documents the user provides.

#### **Main Parts**

#### 1. User Interface

A simple website lets users upload PDFs, choose topics, take quizzes, and ask questions.

## 2. Breaking Down the Documents

When a PDF is uploaded, it is split into smaller overlapping sections to keep the meaning clear.

#### 3. Turning Text Into Searchable Data

These text sections are turned into numbers (called embeddings) that make it possible to find similar content later.

## 4. Storing the Data

The embeddings are saved in a special database that can quickly find similar ones when needed.

#### 5. Finding Topics

The system looks at the uploaded document and picks out the main subjects and related areas. Users can then pick what they want to focus on.

#### 6. Answering Questions

When a user asks something, the system searches for related content from the uploaded document. It also considers what the user and system talked about earlier.

#### 7. Writing the Response

The answer is based only on what's in the document and the ongoing conversation. If the system can't find an answer, it will say so and suggest other areas to look into.

#### 8. Making Quizzes

The system can create short quizzes (3 questions per topic) based on the chosen subject. It removes things like author notes and focuses on the main content.

#### 9. Scoring and Feedback

Quiz answers are checked automatically. Users get a score and see how they're doing over time. The system recommends what to review next.

#### **How It Works?**

## For Questions:

- 1. Looks at current and previous user questions.
- 2. Searches for 3 matching sections from the document.

3. Uses these sections and the conversation to form an answer.

#### For Quizzes:

- 1. Picks 5 relevant content sections.
- 2. Cleans the content to keep only useful material.
- 3. Writes 3 multiple-choice questions.
- 4. Scores the answers and tracks how the user performs.

## **Scoring System**

80% or more: Mastered

60-79%: Clear understanding

Below 60%: Needs review

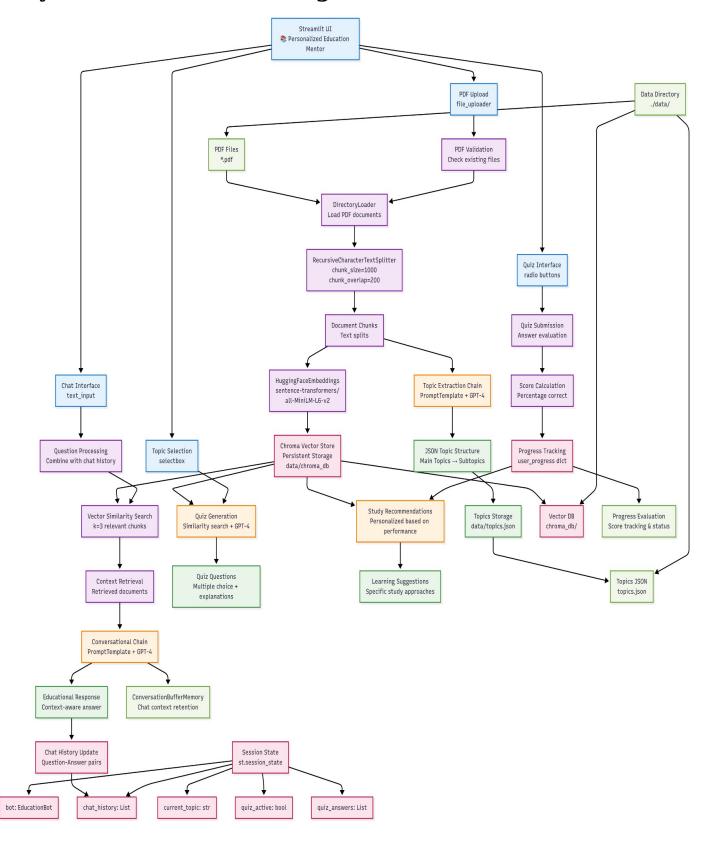
## **Personal Progress**

Tracks how many times a topic was studied and the average score. Offers suggestions based on past performance.

#### Note

This system doesn't add any outside information. All answers and quizzes are based only on the uploaded PDFs and what's been discussed in the conversation.

## **System Architecture Diagram**



## **Discussions**

To create a customized study aid, we integrated semantic search, large language models, and an interactive web interface. Here is what we discovered about what succeeded, what failed, and our next steps.

#### What was Effective?

- **1. Lively Interactions :** Follow-up questions appear natural since the system remembers previous queries and finds pertinent content, saving you from having to start over every time.
- **2. Instant, On-Demand Tests:** Students can attempt practice quizzes with immediate feedback after reading.

## What failed?

- 1. The Detection of Subjects Is Not Perfect: Papers with unusual layouts or nonstandard titles result in messy subject lists that look like rough outlines.
- **2. Quiz Depth Vary :** Unstructured source information often produces surface-level inquiries rather than profound ones.
- **3. Computational Overhead :** Standard laptops are strained by large text processing, and when usage increases, they drastically slow down.

## **Ideas for Future Projects**

- **1. Offer Extra Format Assistance :** The addition of EPUB, DOCX, and HTML would allow students to use e-books, web articles, and other resources directly.
- **2. Intelligent Quiz Level :** After incorrect replies, questions should become easier, and as students demonstrate proficiency, they should become harder.
- **3. Contextual Link-Out:** Direct access to scholarly publications, Wikipedia, or MOOCs for students who wish to go further into a subject.
- **4. Instructor & Parent Views :** Dashboards for teachers and parents to track learning patterns, see class progress and provide targeted support.

## **Conclusion**

This project puts students in control of their learning through an AI-powered study companion. By working with materials students already have, using advanced text processing, and creating interactive practice sessions, we've built a system that can adapt to different learners without requiring massive infrastructure.

Since we built it in separate pieces, adding new features or tweaking it for other subjects becomes straightforward. This project demonstrates that students don't need expensive tools to get personalized help—they just need better ways to work with what they already have.

## References

- 1. Bertsekas, D. P., & Tsitsiklis, J. N. (2002). Introduction to Probability. Athena Scientific.
- 2. LangChain Docs. <a href="https://python.langchain.com/docs/">https://python.langchain.com/docs/</a>
- 3. Streamlit Docs. <a href="https://docs.streamlit.io/">https://docs.streamlit.io/</a>
- 4. OpenAI API Docs. <a href="https://platform.openai.com/docs/api-reference">https://platform.openai.com/docs/api-reference</a>
- 5. Chroma DB Docs. <a href="https://docs.trychroma.com/docs">https://docs.trychroma.com/docs</a>
- 6. HuggingFace Transformers.

https://huggingface.co/docs/transformers/en/index