productivity
project manager / syllabus thing / task manager
Dream Job Tracker with ML Assistant
Adaptive Pomodoro Timer (Instead of the standard 25 minutes of work followed by a 5-minute break, the app would use ML to predict how long the user can maintain focus and adjust the work and break intervals accordingly)
health
alzheimer's detection (through speech or else)
sleep cycle detector and alarm
ingredient checker
Education

Stock portfolio for beginners
Scam detector
Harmful brain activity: using data when someone is relaxed vs data when some brain waves change when getting exposed to posts/videos that affect our brain negatively.

Top 15 ideas:

1. **Personalized education prototype** **Najma**
    a. ***Description:*** This is expanding on Najma's idea to have some kind of road map. We could take in the educational goals of a person (or career goals), current background, and have an AI create a program that leads them through the process. The difference is that this application could keep track of the user's progress so that they can take off right where they left off and more importantly be dynamic. If the user learns in a particular way or has certain learning preferences (more project based learning or passive, etc), it can adjust its teaching style (so this could be some sort of reinforcement learning.
        i. A way it might learn someone's learning style: If it explains a certain concept with some classic/traditional explanation, and the user asks certain questions about the explanation or ask for an explanation from a more mathematical/story-telling point of view, the AI could use this info to inform future explanations. Same thing goes for recommended links or projects. Additionally, we could have an easy rating system for how helpful explanations, projects or articles are for the user.
    b. ***Pros***: Would be a cool and useful tool
    c. ***Cons***: Has several layers of complexity, but we could also prototype only certain parts (potentially)
2. **Alzheimer's/dementia detection (possibly through speech, or otherwise MRIs)**
    a. ***Description:*** Given data on someone's speech, we could see if we are able to detect whether they have dementia within a certain probability (ideally with a microphone with the quality of an iPhone). This would be really cool because it could be used by any person from their home. For more information on the specifics: Context is not key: Detecting Alzheimer's disease with both classical and transformer-based neural language models - ScienceDirect

b. Potential datasets:
    i. [LOGOS Speech Recognition Challenge (USYD) | Kaggle]
c. ***Pros***: Very related to machine learning, a good project to learn more about supervised learning
d. ***Cons***: people are actively doing research on this, so it might be hard to find novelty (although not impossible)

3. **Obsidian plug-in** Najma Elorie
    a. ***Description:*** Obsidian is an open-source note-taking application that is able to make a visual representation of notes through a network as shown in the image below. Several developers have created plug-ins for the platform like a plug-in with github or a plug-in with an apple pencil. We could create a plug-in that uses machine learning to summarize the notes / create a flow chart at the end of a note / that gives suggestions to what other notes could be connected to your current note / can clean up your notes (and the connections between them)

    b. ***Pros***: Would be a useful productivity tool, very hands-on
    c. ***Cons***: Might require us to learn a lot about the obsidian API and the logistics of plug-ins, which might take away from the time we spend focusing on the machine learning aspect

4. **AI productivity tracker**
    a. ***Description:*** This would be able to detect what websites you are using and use them to predict what tasks you are working on from your to do list and how productive you are.

5. ***Investment helper*** Elorie
    a. ***Description:*** Stock Prediction System using RL (implementing Q-learning to predict short term patterns like troughs, etc. - https://github.com/edwardhdlu/q-trader/tree/master/data), or portfolio builder (considers personal values, risk adversity, etc.). Could also make this targetted to beginners (because there are probably lots of expensive applications that probably do this for investment professionals).
    b. ***Potential Datasets:***
        i. GSPC stock trading dataset collected from Yahoo! Finance (ranaroussi/yfinance: Download market data from Yahoo! Finance's API) for short term predictions
    c. ***Pros***: Lots of data (can pick from several types of data – historical stock price, current financial statements, etc), very useful for our lives :), we could also integrate this into tool

d. ***Cons***: these types of projects are common, but we could have a different/novel approach to it (by combining datasets in new ways), there is so much data so there is room for innovation (could involve web scraping which is something I want to learn)

6. ***LMSYS - Chatbot Arena Human Preference Prediction***
Descriptions: This competition challenges you to predict which responses users will prefer in a head-to-head battle between chatbots powered by large language models (LLMs). You'll be given a dataset of conversations from the Chatbot Arena, where different LLMs generate answers to user prompts. By developing a winning machine learning model, you'll help improve how chatbots interact with humans and ensure they better align with human preferences.
Dataset: The competition dataset consists of user interactions from the ChatBot Arena (https://lmarena.ai/).
Features user prompts, chatbot responses, and preference rankings.
https://github.com/anthropics/hh-rlhf
Pros:
   - Existing frameworks like Hugging Face make it easier to preprocess data, train models, and evaluate performance.
   - Tackles a cutting-edge problem in AI and NLP, aligning with the growing interest in improving LLMs' interaction quality.
   - Directly applicable to industries focused on conversational AI (chatbots, virtual assistants).

Cons:
   - Dataset may have inherent biases, such as:
      - Position bias: Users preferring the first response.
      - Verbosity bias: Favoring longer responses.
      - Addressing these biases requires careful feature engineering and testing.
   - It is also subjective to each individual, and might require advance computation

7. ***Child Mind Institute — Problematic Internet Use***
Description: The goal is to develop a predictive model that analyzes children's physical activity and fitness data to identify early signs of problematic internet use. This predictive tool can trigger interventions promoting healthier digital habits. The model could use physical metrics such as exercise frequency, screen time correlation, posture data, and sleep patterns as indicators.
Datasets:
   - The Healthy Brain Network (HBN) dataset
   - **Youth Risk Behavior Surveillance System (YRBSS)**: Tracks health behaviors among youth, including physical activity and screen time.
   - **NIH Adolescent Brain Cognitive Development (ABCD) Study**: Includes behavioral and physical activity data alongside internet use statistics.
   -

Pros
Social Impact: Helps address a critical issue of digital addiction in children, promoting healthier habits.

Feasibility: Physical activity data is widely available from wearables and surveys.
Interdisciplinary: Combines machine learning, health, and psychology for real-world relevance.
Actionable: Provides direct feedback for parents, schools, or health professionals.
Cons
Subjectivity: Problematic internet use is hard to define and varies across individuals.
Bias and Confounding: Other factors (e.g., family environment) might influence results.
Privacy Concerns: Handling sensitive data for children requires extra precautions.
Data Availability: Labeled datasets linking physical activity to internet use may be limited.

8. **AI-Powered Personal Assistant for Task Prioritization** Elorie
Description: Develop an AI assistant that reviews your daily tasks, predicts priorities based on deadlines and importance, and helps you plan your day efficiently.
Works in the following way:
   - Input tasks with deadlines and priorities.
   - Use machine learning to rank tasks based on urgency and effort.
   - Provide reminders or break large tasks into smaller, manageable chunks.

Dataset: Create synthetic task data or use Task Management Logs.
Taks completion dataset : Includes task completion records with metadata like priority, deadlines, and status.
Academic timetable: The dataset contains the academic timetable for each week in each term during a whole academic year at London South Bank University.
Pros:
   - We can Organize our day effectively, Avoid missing important deadlines.
   - This can be used also we upload our syllabus to this platform, it puts all the dates to calendar and prioritise assignments with their weights

Cons:
   - Adapting to user feedback (e.g., changing task priority dynamically).

9. **AI-Powered School Project Planner** Najma
Description: Develop a tool where students can upload their project description or syllabus, and the AI generates an organized project outline, prioritizes tasks, and suggests deadlines. This tool can help students break large projects into manageable steps and stay on track. Provide a Gantt chart or timeline visualization for tracking progress.
Datasets:
Synthetic Data
Project Management LLM Dataset: Descaription: A synthetic dataset representing project performance data, including task descriptions, cost variance, and schedule variance.
Goal and Plan Recognition Dataset: Datasets designed for evaluating goal and plan recognition approaches, containing planning problems with tasks and objectives.
Pros:
Practical Application: Helps students organize their workload effectively.

Customizable: Adapts to different project types (e.g., essays, presentations, group projects).
Time Management: Encourages timely submissions by breaking tasks into smaller steps.
Cons:
Ambiguity in Descriptions: Complex project descriptions may require additional user input for clarification.
Dependencies: Requires accurate recognition of task dependencies for prioritization.

An example of this platform can be this:
https://galaxy.ai/ai-project-plan-generator what is missing here is a calendar view, steps to get started, we can make it personalised by inputting our skills and education, how many people need to work on a project, who does who based on skills and experiences.

10. **AI-Driven Exam Preparation Optimizer** **Najma** **Elorie**
Description: An AI tool that analyzes your syllabus, notes, and past exam patterns to create a customized study plan for your exams. It prioritizes topics based on importance, past difficulty levels, and your personal performance trends.
Input:
Upload your syllabus, lecture notes, and past papers.
Optionally, input your current understanding of each topic (e.g., rate confidence level from 1-5).
Processing:
NLP analyzes syllabus and past papers to identify high-priority topics.
ML ranks topics based on frequency in past exams and syllabus weight.
AI generates a study schedule, assigning more time to weaker topics.
Output:
A visual timeline with prioritized topics, recommended study hours, and break reminders.
Suggestions for extra resources (e.g., videos, tutorials) for weaker areas.
Datasets:
Synthetic data
Riiid Answer Correctness Database: Contains: Over 100 million student interactions, including questions and answers. Can be used to Predict performance trends and topic priorities.
SQuAD (Stanford Question Answering Dataset): Contains: Paragraphs and questions with answers. Use: Train an NLP model to identify important sections in lecture notes or past papers.
EdNet Dataset: Contains: Logs of students' interactions with educational content. Use: Analyze learning behaviors and predict weak areas based on interactions.
ASSISTments Dataset: Contains: Student performance on problems categorized by topic. Use: Rank topics by difficulty and focus on weak areas.

Pros:

1. Tailored to your specific needs and syllabus.
2. Optimizes time by focusing on critical topics first.

Cons:

1. Requires preprocessing to standardize uploaded materials.
2. Limited without access to past papers for pattern analysis.

Extensions:

- Add gamification (e.g., streaks, rewards for completing goals).
- Integrate with apps like Notion or Google Calendar for reminders.

We can also use Pre trained models:

For NLP Tasks

1. **Hugging Face Transformers**
    - Models:
        - BERT (for understanding syllabus and lecture notes).
        - T5 or GPT-3 (for generating study plans and summaries).
    - Use: Analyze text data (e.g., notes, papers) and extract important topics.
2. **spaCy**
    - Use: Quickly preprocess and extract named entities (e.g., topic names) from lecture notes or syllabi.

For Text Summarization

- **BART (Hugging Face)**
    - Use: Summarize lengthy notes into concise points for study plans.

For Recommendation Systems

- **LensKit**
    - Use: Create a system that suggests additional resources based on weak topics.

FINAL IDEA:

- A model that takes lecture notes, syllabus, exam time table, and information about the user's studying habits to generate a customized study plan for your exams
- Text-based model
- Website
- Login feature & account info (from past sessions)
- Course tabs after login with status of progress and history of conversation
- Not saving a file in our cloud, deleting the files after use log out
- Save button to save info generated for users
- The model generated text-based timeline customised to each user
- GUI if not website

Future works:

- Possibly generate cheatsheet
- Integration with Notion
- Publishing the paper
- Interpreting images and generating texts

Resources:

Machine Learning-Based Automatic Timetable Generation | Journal of Soft Computing and Computational Intelligence (e-ISSN: 3048-6610) (matjournals.net)

Datasets:

Riiid Answer Correctness Database: The database contains several CSV files, including train.csv, test.csv, lectures.csv, questions.csv, which contains 101 million entries and data on answer correctness, user performance, question difficulty, and lecture content, useful for predicting future performance and creating a customized study plan.

Open University Learning Analytics Dataset: This dataset offers two of the elements, behavior and performance and has data about 22 courses, 32,593 students, their assessment results, and logs of their interactions with the VLE represented by daily summaries of student clicks (10,655,280 entries).
EdNet Dataset: EdNet dataset contains various features of student actions such as which learning material he has consumed, response, how much time he has spent solving a given question or reading through expert's commentary.

We will focus on the Open University Learning Analytics Dataset because it balances data on student performance, engagement with lectures, and interaction logs, which makes it a better choice for understanding how students interact with courses and to predict study patterns better.
MOOC Dataset: The dataset with over 3,000 entries has data on student demographics, course details, engagement, progress, and assessment scores, tracking actions like video views and quiz participation which could be useful for suggesting a study plan and performance evaluation.

https://galaxy.ai/ai-project-plan-generator

[Open University Learning Analytics Dataset:](#) This dataset offers two of the elements, behavior and performance and has data about 22 courses, 32,593 students, their assessment results, and logs of their interactions with the VLE represented by daily summaries of student clicks (10,655,280 entries).

Example

An example of a single piece of synthetic data generated by GPT model o1:

Note: ChatGPT was used to generate these tables.

Name	Study Hours/Day	Confidence Level	Study Pattern	Preferred Study Time	Exam Name	Exam Date	Exam Time	Duration	Subject	Topics	Weightage	Notes
John Doe	4	Medium	short_sessions	evening	Mathematics 101	2025-02-15	10:00	2 hours	Mathematics 101	1. Algebra 2. Calculus 3. Differential Equations 4. Linear Algebra	Algebra: 25 Calculus: 40 Differential Equations: 20 Linear Algebra: 15	• Lecture 1: Introduction to Algebra • Lecture 2: Limits and Continuity • Lecture 3: Integration Techniques • Lecture 4: Eigenvalues and Eigenvectors

code_module	code_presentation	id_assessment	assessment_type	date	weight	module_presentation_length	activity_type	week_from	week_to	id_student	gender	region
AAA	2013J	1752	TMA	19	10	268	oucontent	2	2	11391	M	East Anglian Region
highest_education	**imd_band**	**age_band**	**num_of_prev_attempts**	**studied_credits**	**disability**	**final_result**	**date_submitted**	**is_banked**	**score**	**id_student**	**id_site**	**sum_click**
HE Qualification	90-100%	55<=	0	240	N	Pass	18	0	78	28400	546652	4

- Knowledge graphs
- education for ai workshop
- Can create synthetic data using GPT but need to check quality