VGC Group Project Proposal

Tips: We strongly recommend you start early to discuss your proposal with your teammate to make sure you are on track to meet the assignments' requirements early on.

Title:
Early identification of distress signals for students for the promotion of mental health using classification of forests.

1. **(5pts) Honor Code and LLM Usage for this Proposal:**
   The usage of large language models into making this proposal was minimal, only being used in order to improve legibility of the text and grammar

2. **(10pts) Learning Objectives:**
   List at least two learning objectives for your project. These should include both conceptual understanding and skill development. This could be based on ideas from earlier in the course or if there is something you've always wanted to try or learn more about.
   - **Conceptual Understanding:** Learn about random forest classification as a classic machine learning algorithm
   - **Conceptual Understanding:** Understand the inference and predictability of machine learning models.
   - **Conceptual Understanding:** Understand the impact of the dataset properties into the problem resolution.
   - **Skill development:** Develop proficiency in the application of machine learning algorithms into real world datasets.
   - **Skill development:** Acquire experience working with machine learning libraries such as scikit learn in Python.
   - **Skill development:** Achieve the ability to critically evaluate the performance of a machine learning model
   - **Skill development:** Gain insight into the process of passing from code to a working website.

3. **(10pts) Timeline:**
   Outline how you plan to spend 5-10 hours on your project. Break down the time into specific tasks or milestones. Here is an adjustable schedule to get you started. Details should be 50-100 words each.

Time	**Task**	**Details**
Hour 1-2	Research and gather resources	Reviewing classification trees and random forest slides to understand key concepts. Explore tutorials and templates for building websites, focusing on AI project deployment. Collect relevant research papers to understand best practices and relevant developments. Identify datasets and external resources necessary for the project implementation. This stage ensures a solid foundation and references materials to guide coding, data analysis, and deployment efficiently.
Hour 3-4	Design the project structure and plan	Plan the project workflow, including data, analysis, model training, and evaluation. Outline milestones for each step to maintain clear progress. Draft an AI deployment strategy for website, considering backend and frontend integration. Decide on tools and frameworks such as Python with Scikit-learn, Node.js and React. This planning sets a structured roadmap to reduce errors and improve efficiency during coding and implementation.
Hour 5-6	Start coding the basic functionalities	Building the main project features. Write initial scripts to explore the data. Train basic models using Scikit-learn, making sure the results are easy to understand. Create a GitHub repository and set

Time	Task	Details
		up Node.js React for the web interface. Focus on making a working example, even very simple, to serve as a base for adding more features later.
Hour 7-8	Test and debug the initial version	Improve the website by adding visuals like graphs and charts. Show how well the model works and presents the results clearly on the site. Connect the frontend and backends smoothly and include AI deployment features. Use user feedback to make the models and code better. This step makes the project easier to use and more functional.
Hour 9-10	Refine and add advanced features	Test the website and AI functionalities for bugs and performance issues. Document code and results clearly. Ensure that GitHub repository is organized and includes instructions for reproduction. Make final adjustments to visualizations, layout, and deployment scripts. This final stage prepares the project for submission.

4. **(15pts) Final Product Description:**
   Describe the final product at three levels**. Mention key features, target audience, user stories, platforms, or any other technical details necessary** (25-100 words).

I. **Minimum Viable Product (MVP):**
   A basic version of your project that meets the essential requirements.
   1. The most basic version of the project will be a simple website (hosted by a third party service) offering a short survey for students of different

ages. Once finished, the website will show a simple true – false result showing if the user is suffering depression or not based on the dataset, and where to find help if its needed

II. **Target Product:**
The version you aim to complete within the given timeframe.
1. For this version we aim to achieve a prettified user-friendly interface, offering visitors an extended survey. The questions will be classified into different sections, and transitions between sections will be added. Once finished, the website will show the probability of suffering depression, anxiety, or mental problems based on the answers. A chart explaining the relation of the user in relation to the data will also be included

III. Reach Version:
1. This version aims to offer personalized guidance based on the survey using openai API models and guardrails. The visitors will be able to respond to multiple surveys of their choice, targeted to different audiences and conditions. Upon completion, they will have a deep dive into the data of similar profiles thanks to the dataset usage and RAG reinforced LLM models.

5. **(10pts) Consultation and Use of LLMs:**
Each student must create a unique project but is allowed to consult with other people and use Large Language Models (LLMs). Address how you plan to incorporate these resources into your project:
We plan to use LLMs as an assistant in programming; however, it is not our first choice while developing this project. We hope to use resources from classes and the internet when implementing the machine learning algorithm. We may ask for LLM assistance when creating the interactive webpage for the project. We do not have much experience with Node.js or deploying websites. If we consult LLMs, we will document their use.

- **Consultation Plan:**
Describe how you expect to seek advice or feedback from peers, mentors, or online communities (25 to 100 words).
    - We plan to first seek advice from each other. We all have different skillsets, so we want to utilize everyone's knowledge. Second, we can seek advice and feedback from our instructors or teaching assistants. This can be through course resources or one-on-one interactions. We may also consult online communities if we are not able to get our questions answered.

- **Use of LLMs:**
  Explain how you will utilize LLMs to assist with coding, debugging, or generating ideas (25 to 100 words). List a minimum of 3 related online sources. Your project should not be a copy of an existing online project, it needs to be a variation or expansion, and the proposal should make this clear.
    - We may utilize LLMs when debugging our machine learning algorithm. This will be if we cannot figure out the problem(s) after consulting our peers or instructors/ teaching assistants.
    - We may utilize LLMs while coding the webpage or generating ideas for how to implement it. We do not have as much experience with this side of the project.
    - Sources: https://nodejs.org/en/learn/getting-started/introduction-to-nodejs , https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html , https://www.geeksforgeeks.org/dsa/random-forest-classifier-using-scikit-learn/
    - The resources linked above include examples for how to implement the machine learning algorithm, but we are using a different dataset with a different classification goal. In addition, we are creating an interactive website that users can access to get predictions based on the trained model. This is a step further from the algorithm simply implemented in code.