BAX 452 Project Proposal - Group 34
Members: Anakin Liu, Ruiyuan Yang, Jerry Xia

In this project, we will utilize a variety of machine learning techniques to help doctors and neural scientists assess Alzheimer's risks.

1. Topic and Dataset

The major aim of this project is to provide a preliminary Alzheimer risk prediction based on a variety of easily attainable biographical information and some commonly available biological metrics of a patient, without having to perform in-depth neurological exams.

Dataset: Alzheimer's Prediction Dataset (Global) - https://www.kaggle.com/datasets/ankushpanday1/alzheimers-prediction-dataset-global
Containing **1** label, **23** features from **74000+** samples

2. Models:

a. **Decision Tree-Based Models (Random Forest, Boosting)** – Primary classification algorithms for capturing complex patterns and interactions in the data.
b. **Lasso Regression** – Used for potential feature selection to simplify the model.
c. **Logistic Regression** – Serves as a baseline control model.

3. Evaluation metrics:

- **AUC-ROC Score (Area Under the Receiver Operating Characteristic Curve)** - Measures the model's ability to distinguish between Alzheimer's and non-Alzheimer's cases.
- **Accuracy** - Evaluates overall classification correctness.
- **Precision-Recall Curve (PRC)** - Provides insights into the model's effectiveness in identifying positive cases.