Course: *Human Centered Data Science*
Team: *Saad Waseem, Yasemin Mutlugil, Nino Sabella, Orkun Akyol, Jingren Dai*
Instructor: *Claudia Müller-Birn*
Date: *July 15th*

Remedi: An Alzheimer's Risk Assessment Tool

Introduction

This project consisted of creating a **Streamlit application** that serves as an interface for users to interact with, and the topic selected by our team was **Alzheimer's disease diagnosis** using AI. In terms of ethical considerations, the models used to diagnose patients should not have biases that discriminate against certain groups of people. This is a common problem with such models. For example, features like gender, age, and education level should not be subjects of discrimination, yet they often are. Our goal in this project was to demonstrate, using explainability methods, what features were being prioritized in model predictions and if they pushed the model towards declaring Alzheimer's or away from making the diagnosis.

Project Goals

The three pillars of our project in terms of goals were stakeholders, explainability, and precision.

The first goal, based on stakeholders, was primarily about understanding our user. In our case, the primary stakeholder was the physician, so we conducted research on what a physician does, why they do what they do, and how the physician goes about making that critical Alzheimer's diagnosis that the model serves for. At the end of the day, the physician is the person making the decisions and assuming the accountability. This is deeply rooted in the ethical need for human oversight when interacting with AI, as stated in the AI Act. Our application is merely a tool that helps the physician make better decisions with as much information as possible.

The second goal, about explainability, is an extension of the first goal. It takes the first goal and pushes it a little further. To assume accountability as the physician, they should know "the why", what made the model recommend diagnosing the patient with Alzheimer's or not. It was immediately clear to us that our application needed to have explanations of what features were most important in the decision and how. This was done through easily interpretable and aesthetically pleasing graphs on the application's prediction page.

The third and final goal was about precision. The application needs to work. The models need to be reliable. Even if the model has weak spots, it should be smart enough to identify when predictions are less or more likely to be right. To build trust with the user (physician), the model should return results with confidence about the diagnosis made based on the inputs given.

Intended Target Audience

The primary user of our tool is a primary care physician who is someone with clinical expertise in diagnosing conditions like Alzheimer's, but likely with only basic familiarity with machine learning. This tool is designed to support, not replace, their medical judgment. It provides an additional layer of insight by highlighting the likelihood of Alzheimer's based on patient data and clearly showing which factors contributed most to the prediction. With this, physicians can better understand the underlying reasoning of the model, validate it against their own expertise, and make more informed decisions with confidence.

Dataset and Documentation

The dataset used in the creation of our models is called "*Alzheimer's Disease Dataset*"[1], and it consists of 32 features, 2149 instances (patients). The data is shared by Rabie El Kharoua on Kaggle.

The dataset is synthetic and made for educational use. The dataset consists of some interesting features such as demographics, lifestyle habits, medical history, clinical metrics, and cognitive and functional evaluations. The target label is a binary (1 or 0) diagnosis of Alzheimer's for each patient (row). 1 means "Has Alzheimer's" and 0 means "No Alzheimer's".

Key Design Decisions

The design of our application was guided by the goal of supporting **physicians in assessing the risk of Alzheimer's Disease** in patients. While the app provides AI-driven predictions, it is built with the understanding that final decisions should always lie with medical professionals. A core focus throughout development was ensuring **interpretability and trust**, so that physicians can understand, interrogate, and evaluate the model's predictions with clarity and confidence.

[1] Dataset: https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset/data

Our development process included multiple rounds of iteration informed by **user interviews**. A Figma prototype was used to test interaction flow and early visual ideas, and the design was refined based on qualitative feedback, especially through **think-aloud evaluations**. The final version of the application reflects both technical feedback and the practical needs of healthcare professionals.

To enable users to easily and intuitively navigate the app, the following pages were included:

- **Dashboard**: A central hub providing an overview of the tool's purpose and structure.

- **Prediction Page**: Allows physicians to enter patient data and receive an Alzheimer's risk prediction with detailed interpretation from multiple models.

- **Methodology Page**: Explains model training, evaluation metrics, and fairness considerations.

- **Resources Page**: Offers helpful materials to further support physicians and explain clinical relevance.

The following key design decisions were made during development:

Intuitive Navigation

The side menu allows for easy access to all sections, and prompts have been added to guide users to supporting content.

Improved Input Design

To accommodate a range of users, **medical jargon** was either removed or accompanied by tooltips. **Proper input controls** (like sliders and drop-downs) were refined based on best UI practices.

Simplified Model Selection

Instead of burdening users with choosing a model without context, we now **display results from all models simultaneously**, allowing for direct comparison. Information about trade-offs between accuracy and interpretability is placed nearby.

Explainability-Centered Visuals

We integrated **SHAP force plots, waterfall charts, partial dependence plots, and violin plots** to give both global and local explanations.

Clear Recommendations

While the app does not prescribe action, it now includes **suggested next steps** and clarifying notes on how to interpret the predictions. The **methodology page** also includes information on data sources, model training, and fairness to help physicians assess the model's **trustworthiness**.

This structured and feedback-driven approach ensures that our application supports **clinical decision-making without overstepping its role**, giving physicians the clarity and confidence needed to assess Alzheimer's risk with the help of machine learning.

Challenges

Our team of five came from diverse backgrounds, but none of us had prior experience designing user interfaces, especially not for explainable AI tools. Streamlit, the framework we chose, was also new to us. Much of the early development involved self-directed learning through documentation and experimentation, which added time and complexity but ultimately improved our technical skills.

We also faced conceptual challenges, particularly when it came to balancing **model performance** with **interpretability**. In several cases, we had to deliberately choose simpler or less optimized techniques to maintain transparency. For example, we removed feature scaling because it distorted the interpretability of SHAP values, and we avoided dimensionality reduction techniques like PCA that obscure individual feature contributions. These tradeoffs shaped both the design and the philosophy behind our tool.

Another key challenge was ensuring fair and balanced data representation. Since our dataset had class imbalances, we made adjustments to ensure that all classes were equally represented during training. This improved both the reliability of predictions and the fairness of the model's behavior. While these decisions weren't always easy, they reflect our commitment to building a tool that is not only technically sound but also trustworthy and user-centered.

Reflections

This project was a valuable learning experience in **human-centered data science**. We realized that building accurate models is only one part of the process. What truly matters is delivering a tool that serves the needs of the primary user. In our case, that user is the physician, who should feel empowered to understand, trust, and ultimately make the final decision based on the model's output. That means making the reasoning behind predictions transparent and easy to grasp, without requiring technical expertise.

Our goal was to support physicians by providing access to three interpretable and reliable models through a simple, intuitive interface. With just a few clicks, they can consult different perspectives on Alzheimer's risk while staying firmly in control of the diagnosis process. This not only enhances their confidence but also promotes efficiency in clinical decision-making. On a broader level, we aimed to align our design with the principles of ethical and responsible AI, where human oversight, transparency, and usability are central.