Project 4 Group 9

Luisa Murillo, Monica Mitry, Gorgina Kareem, & Adriana Kuhl

Introduction:

The project team chose the Alzheimer's Disease dataset because it provides a rich set of data related to the early diagnosis and progression of Alzheimer's disease, which is a critical area of healthcare. Alzheimer's disease impacts millions of people globally, and early detection plays a crucial role in providing better patient care and treatment. This dataset offers an opportunity to apply machine learning techniques to predict the likelihood of Alzheimer's disease, which could potentially help improve the accuracy of diagnoses and aid in the development of treatment strategies. Moreover, it aligns with my interests in healthcare data analytics and machine learning.

What was the inspiration?

The inspiration for this project came from the increasing importance of data-driven approaches in the healthcare field, particularly in the diagnosis and management of diseases like Alzheimer's. Alzheimer's disease is a growing concern as the global population ages, and its early diagnosis is critical for improving the quality of life for patients. As healthcare data becomes more accessible, machine learning has proven to be a powerful tool in predicting and understanding complex diseases. The goal of this project is to explore how predictive analytics can assist in identifying potential Alzheimer's patients at an earlier stage, providing them with the best chance for intervention and care.

Dataset Link: 🧠 Alzheimer's Disease Dataset 🧠

High Level Questions:

- Is there an average age that Alzheimer's symptoms begin to appear?
- Is there a correlation between ethnicity and a higher risk of Alzheimer’s?
- Does a patient's lifestyle play a role in the risk of Alzheimer’s?
- Do other health concerns, such as cholesterol or depression, increase the risk of Alzheimer’s?
- What is the most common symptom for Alzheimer’s?

What are we visualizing?

In this project, we are visualizing various aspects of Alzheimer's disease data, focusing on key features that could impact the prediction of the disease. The visualizations in Tableau will include:

1. **Age distribution:** Visualizing with a bar graph the relationship between age and Alzheimer's disease occurrence.
2. **Gender comparison:** Displaying how gender impacts the likelihood of the disease with a pie chart.
3. **Ethnicity and risk factors:** A comparison chart (bubble chart) to determine if there is a correlation between ethnicity and a higher risk of Alzheimer's.
4. **Lifestyle Impact:** Visualizing lifestyle factors with a heatmap (e.g., diet, exercise, smoking) and their relationship to Alzheimer's risk using scatter plots and trend lines.
5. **Feature correlations:** Heatmaps and scatter plots to show relationships between key features like cognitive scores, age, and other health indicators.
6. **Prediction outcomes:** A clear visualization of predicted Alzheimer's disease status (e.g., positive/negative predictions) to understand the model's performance.
7. **Other Health Concerns Correlations:** Heatmaps and correlation charts to analyze whether conditions like cholesterol or depression increase the risk of developing Alzheimer's.
8. **Common Symptoms:** A frequency-based visualization (bar chart or pie chart) highlighting the most common reported symptoms in Alzheimer's patients.

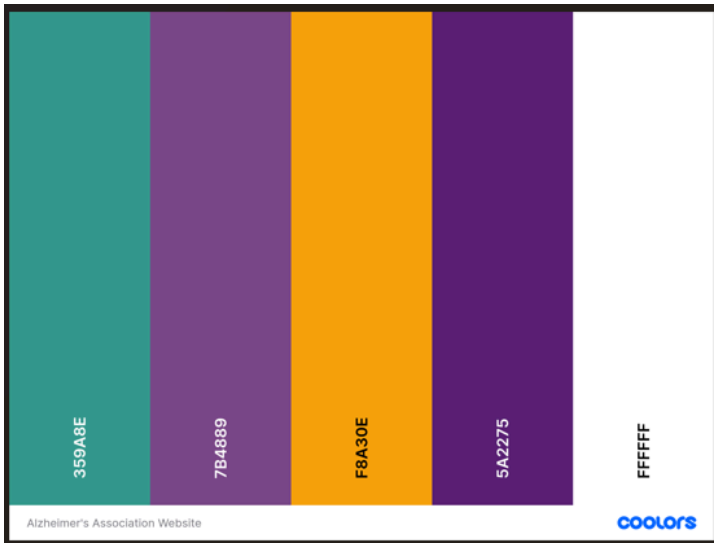
The goal of these visualizations is to provide both a high-level overview and detailed insights into the factors that influence the development of Alzheimer's disease, along with a predictive model that can assist in early diagnosis.

What are you predicting?

We will be using the classification method of machine learning to predict whether or not a person is likely to be diagnosed with Alzheimer's. Our target column in our dataset is labeled: Diagnosis. The algorithms we will be testing against the dataset are: Linear Regression, Decision Trees, Random Forest, and XGBoost.

Color Palette

National Alzheimer's Association branding



Roles & Responsibilities:

- Flask App - Everyone Helps
- Data Cleaning - Everyone Helps
- Tableau - Monica & Luisa
- Machine Learning - Adriana & Gorgina
- Slides - Adriana & Monica
- Write Up - Gorgina & Luisa