# Technical Audit of an Automated Decision System (ADS) Predicting Early Alzheimer Using MRI Data

**Aditya Dixit** AD5157@NYU.EDU
*Courant Institute of Mathematical Sciences*
*New York University*
*New York, NY, USA*

**Ayush Jain** AJ3152@NYU.EDU
*Courant Institute of Mathematical Sciences*
*New York University*
*New York, NY, USA*

**Gurmehr Sohi** GS3541@NYU.EDU
*Courant Institute of Mathematical Sciences*
*New York University*
*New York, NY, USA*

## Abstract

Automated decision-making systems (ADS) are omnipresent in today's healthcare systems. They help in providing accurate, critical, cost and time-efficient solutions to a wide array of medical conditions which led to their widespread adoption. However, it would be naive to underestimate the issues such systems face in a real-world setting. Such systems are prone to various biases and errors. Institutional biases stemming from race, gender, age, income, and ethnicity are endemic in Healthcare. In this study, we are conducting a technical audit of an ADS predicting Early Alzheimer's trained on the OASIS (Open Access of Series of Imaging Studios) dataset having a total of 807 MRI scans of 566 patients across OASIS-1 and OASIS-2 datasets.

**Keywords:** Automated Decision System, Alzheimer, Dementia, Magnetic Resonance Imaging, Open Access of Series of Imaging Studios

**Datasets:** OASIS-1, OASIS-2

## 1. Introduction

Advanced computational models called Automated Decision Systems (ADS) use machine learning, artificial intelligence, and big data analytics to generate forecasts, suggestions, or judgments on their own. ADS has recently attracted a lot of interest and has been widely adopted across a variety of industries thanks to its capacity to analyze enormous amounts of data, spot patterns, and draw insights with little human involvement.

The exponential rise of data produced by electronic health records (EHRs), medical imaging, genomic sequencing, and other sources is one of the main factors behind ADS's success in healthcare. This wealth of data makes it possible to make more accurate and complex ADS models that can help doctors make better decisions. By using advanced techniques like deep learning and natural language processing, ADS can get useful insights from complex and unstructured data. This makes it possible to find secret relationships, trends, and

abnormalities that might not be obvious to human experts.

ADS has a wide range of potential advantages for the healthcare industry, from better diagnostic precision and lower healthcare costs to improved patient care and more effective resource management. For example, ADS can aid in the early diagnosis of conditions including cancer, Alzheimer's disease, and diabetes, enabling prompt treatment and improved patient outcomes. ADS can also provide more individualized treatment regimens and direct preventive steps by foreseeing probable consequences and identifying high-risk patients, ultimately helping to create a healthcare system that is more patient-centric.

The early diagnosis of Alzheimer's disease, a neuro-degenerative condition that affects millions of individuals worldwide, is one of the most promising uses of ADS in the medical field. Alzheimer's disease is a progressive neurodegenerative disorder that mostly affects older people and is the most common cause of dementia in the world. It is characterized by a gradual loss of cognitive skills, such as memory loss, trouble with language, and trouble thinking, which makes it harder to do daily chores and stay independent. Still, no one knows exactly what causes Alzheimer's, and there is no fix for the disease. But early detection and treatment can help slow the progression of Alzheimer's, handle its symptoms, and improve the quality of life for people with the disease and their caregivers. By allowing for prompt therapies and possibly slowing the disease's course, an early diagnosis of Alzheimer's can significantly improve a patient's quality of life.

Even though ADS has many benefits, its growing use of it in healthcare raises some ethical and practical issues. One of the biggest problems is making sure that the ADS models are fair, open, and accountable. Sometimes they act like black boxes, making it hard to figure out how they work and how they make decisions. It is very important to deal with problems of bias, discrimination, and fairness so that existing differences don't get worse and so that everyone can get the benefits of ADS.

Also, the performance of ADS needs to be rigorously validated, tested, and constantly monitored to keep it accurate, reliable, and safe. Another important part is making sure that sensitive patient data is safe and private since ADS needs a lot of personal information to work well. To solve these problems and prepare the way for the responsible use of ADS in healthcare, it is important for all stakeholders, such as healthcare workers, data scientists, ethicists, and policymakers, to work together.

## 2. Background

This Automated Decision System (ADS) is intended to estimate a patient's Clinical Dementia Rating (CDR) by using relevant features from MRI datasets given by the Open Access Series of Imaging Studies (OASIS) project. The ADS uses powerful machine learning and data analytics to look for early signs of Alzheimer's disease and figure out how bad a person's cognitive impairment is. By accurately predicting CDR scores, the system helps doctors and nurses make better decisions about how to diagnose, treat, and take care of conditions linked to dementia. Also, the ADS makes an early intervention, risk stratification, and

constant monitoring of the progression of the disease possible. This makes Alzheimer's care more patient-centered and individualized.

### 2.1 ADS Purpose and Goals

The main goal of this Automated Decision System (ADS) is to use cross-sectional and longitudinal MRI data to predict the start of dementia, especially Alzheimer's disease.

The following are its stated goals:

1. **Early detection:** If the ADS can correctly predict dementia in its early stages, healthcare providers will be able to start the right treatments, which will help patients and the public's health as a whole. Improve the accuracy of the diagnosis: The ADS tries to reduce the chance that a person will make a mistake when detecting dementia by using advanced machine learning algorithms, which may give more accurate and consistent results.

2. **Increasing efficiency:** The ADS aims to save healthcare workers time and money by automating the decision-making process. This will let them focus on giving their patients high-quality care.

3. **Monitoring disease progression:** By keeping track of how CDR scores change over time, the ADS hopes to be able to track how a disease is getting worse and help clinicians figure out how effective interventions are, change treatment plans, and make smart choices about patient care.

4. **Supporting clinical decision-making:** The ADS is an extra tool for health care professionals. It gives them valuable information about the mental health of their patients and helps them make more accurate and well-informed decisions about diagnosing, treating, and managing conditions linked to dementia.

### 2.2 Trade-off among Multiple Goals

While the ADS is designed to achieve multiple goals, such as early detection, increasing efficiency, Monitoring disease progression, and supporting clinical decision-making, it is possible that trade-offs may arise between these objectives. Some potential trade-offs include:

1. **Fairness vs. accuracy:** Fairness in the ADS predictions for different subgroups can call for certain accuracy compromises, especially if specific clinical or demographic characteristics have a strong correlation with the outcome variable (CDR). To be fair, the system may need to account for possible biases and change its predictions accordingly, which could hurt its overall performance. It's important to find a balance between fairness and accuracy to make sure that the benefits of the ADS are shared fairly among all patient groups.

2. **Complexity vs. interpretability:** The ADS may use complicated machine learning models, such as deep learning or ensemble methods, to improve its ability to predict.

Even though these models can make the system work better, they can also make it harder to understand, making it hard for healthcare workers to figure out why the predictions are what they are. For the ADS to be trusted and to be useful in clinical settings, there needs to be a balance between how complicated the model is and how easy it is to understand.

3. **Sensitivity vs. specificity:** In order to find people with cognitive impairment as soon as possible, the ADS can be tuned to increase its sensitivity, or its ability to correctly identify people with cognitive impairment. But this could make the test less accurate and increase the number of false positives, where people who don't have cognitive decline are wrongly labeled as having it. To make accurate and trustworthy predictions, it is important to find a good balance between sensitivity and specificity.

To deal with these trade-offs, the ADS's goals need to be carefully balanced, and there may be a need for continued improvements and monitoring to make sure that everyone gets the best results.

## 3. Input and Output

### 3.1 Dataset Description and Collection

The data used by this ADS comes from the Open Access Series of Imaging Studies (OASIS) project, which makes neuroimaging datasets freely available to the research community. The OASIS-1 and OASIS-2 datasets are used in the ADS. These are made up of both cross-sectional and longitudinal MRI data.

The OASIS-1 collection is made up of cross-sectional MRI data from a wide range of people, such as young adults, older adults, and people with Alzheimer's disease. The dataset includes a wide range of demographic and clinical factors, giving a snapshot of each participant's cognitive health at a single time point.

On the other hand, the OASIS-2 dataset is made up of longitudinal MRI data collected from participants at various time points. This makes it possible to look at how the structure and function of the brain change over time. This dataset includes individuals with varying degrees of cognitive impairment, from no dementia to severe cognitive impairment, and provides valuable insights into the progression of Alzheimer's disease and other dementia-related conditions.

Both the OASIS-1 and OASIS-2 datasets were collected through a rigorous process that included recruiting participants, getting their informed permission, and following standard protocols for data collection. The MRI data were collected with a number of different MRI scanners and imaging sequences. This made sure that the data were of good quality and consistent. Also, the datasets have been carefully pre-processed, anonymized, and organized so that academics and developers can easily access them and look at them.

By using the OASIS-1 and OASIS-2 datasets, the ADS can take important features from the MRI data to estimate the Clinical Dementia Rating (CDR) and measure the severity of

cognitive impairment in a person. With the help of these large and high-quality datasets, more accurate and reliable models can be made to predict Alzheimer's disease and track its growth.

### 3.2 Dataset Features

The OASIS-1 dataset contains cross-sectional MRI data and relevant demographic and clinical information from a diverse group of participants. While the specific features in your dataset may vary, some common features in the OASIS-1 dataset include:

- ID: A unique identifier assigned to each participant. Datatype: Nominal categorical
- Age: The age of the participant at the time of the MRI scan. Datatype: Continuous numerical
- Gender (M/F): The participant's gender (male or female). Datatype: Binary categorical
- Education (Educ): The participant's level of education (e.g., years of education or educational attainment).
- Socioeconomic status (SES): The participant's socioeconomic status, is often represented as an ordinal variable with lower values indicating higher status.
- Mini-Mental State Examination (MMSE): A cognitive test score used to assess cognitive function, with higher scores indicating better cognitive performance.
- MRI features (eTIV, nWBV, ASF): Various features extracted from the MRI scans, such as volumetric measurements, cortical thickness, or surface area of different brain regions.

The above list provides a general overview of the features commonly found in the OASIS-1 and OASIS-2 datasets.

Also, there are correlation coefficients, like Pearson's or Spearman's rank correlation, that can be used to figure out how two traits are related to each other. These associations help figure out if there might be a link between two or more input features, which can help with feature selection and preprocessing.

### 3.3 Data Profiling

We profiled the dataset formed by concatenating the OASIS-1 and OASIS-2 datasets. The following figures visualize the data profiling results received during the study.

```
M/F       object
Hand      object
Age        int64
Educ     float64
SES      float64
MMSE     float64
CDR      float64
eTIV       int64
nWBV     float64
ASF      float64
dtype: object
```

**Figure 1:** Data Type for each feature

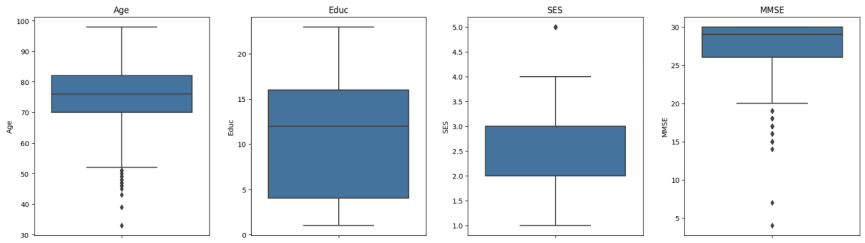


**Figure 2:** Boxplots-1 describing Statistics for each feature

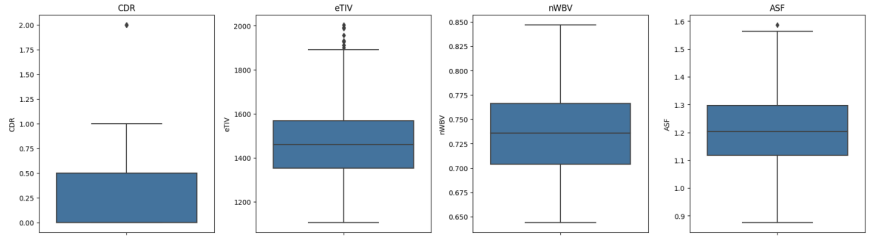


**Figure 3:** Boxplots-2 describing Statistics for each feature

| | Age | Educ | SES | MMSE | CDR | eTIV | nWBV | ASF |
|---|---|---|---|---|---|---|---|---|
| count | 608.000000 | 608.000000 | 570.00000 | 606.000000 | 608.000000 | 608.000000 | 608.00000 | 608.000000 |
| mean | 75.208882 | 10.184211 | 2.47193 | 27.234323 | 0.288651 | 1477.062500 | 0.73713 | 1.203597 |
| std | 9.865026 | 6.058388 | 1.12805 | 3.687980 | 0.377697 | 170.653795 | 0.04267 | 0.135091 |
| min | 33.000000 | 1.000000 | 1.00000 | 4.000000 | 0.000000 | 1106.000000 | 0.64400 | 0.876000 |
| 25% | 70.000000 | 4.000000 | 2.00000 | 26.000000 | 0.000000 | 1352.500000 | 0.70400 | 1.118000 |
| 50% | 76.000000 | 12.000000 | 2.00000 | 29.000000 | 0.000000 | 1460.000000 | 0.73600 | 1.202000 |
| 75% | 82.000000 | 16.000000 | 3.00000 | 30.000000 | 0.500000 | 1569.000000 | 0.76625 | 1.297500 |
| max | 98.000000 | 23.000000 | 5.00000 | 30.000000 | 2.000000 | 2004.000000 | 0.84700 | 1.587000 |

**Figure 4:** Distribution of all features in our Data

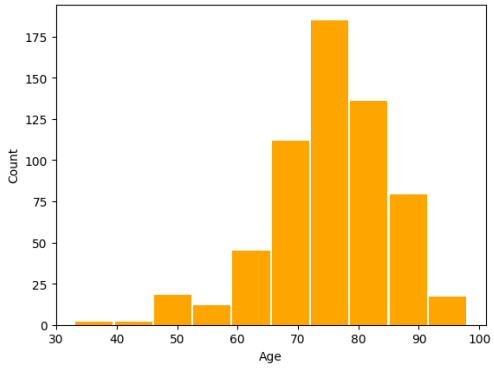


**Figure 5:** Histogram for Age distribution in the Data

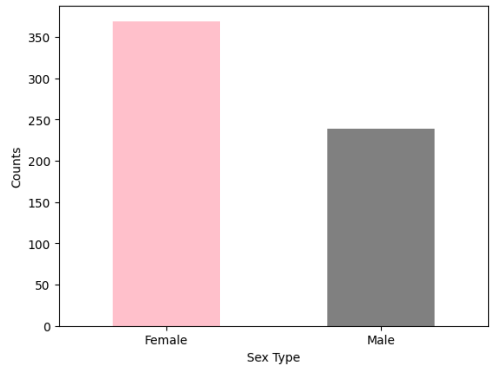


**Figure 6:** Number of Males and Females in our Data

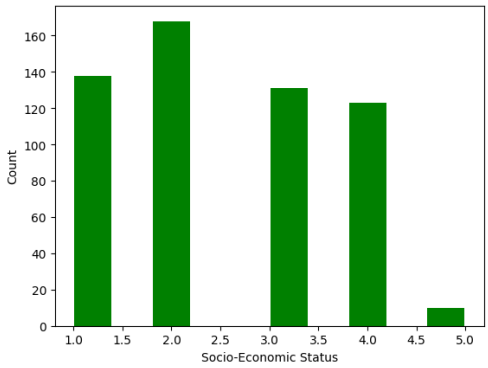


**Figure 7:** Distribution of Socio-economic status in the Data

Our heatmap in figure.8 displays the values of the correlation between the variables of the datasets. Each cell in the heatmap shows the correlation value between the two features. We use Pearson correlation coefficient which gives us a coefficient value. Value -1 represents a perfect negative correlation while value 1 represents a perfect positive correlation and 0 indicates no correlation.

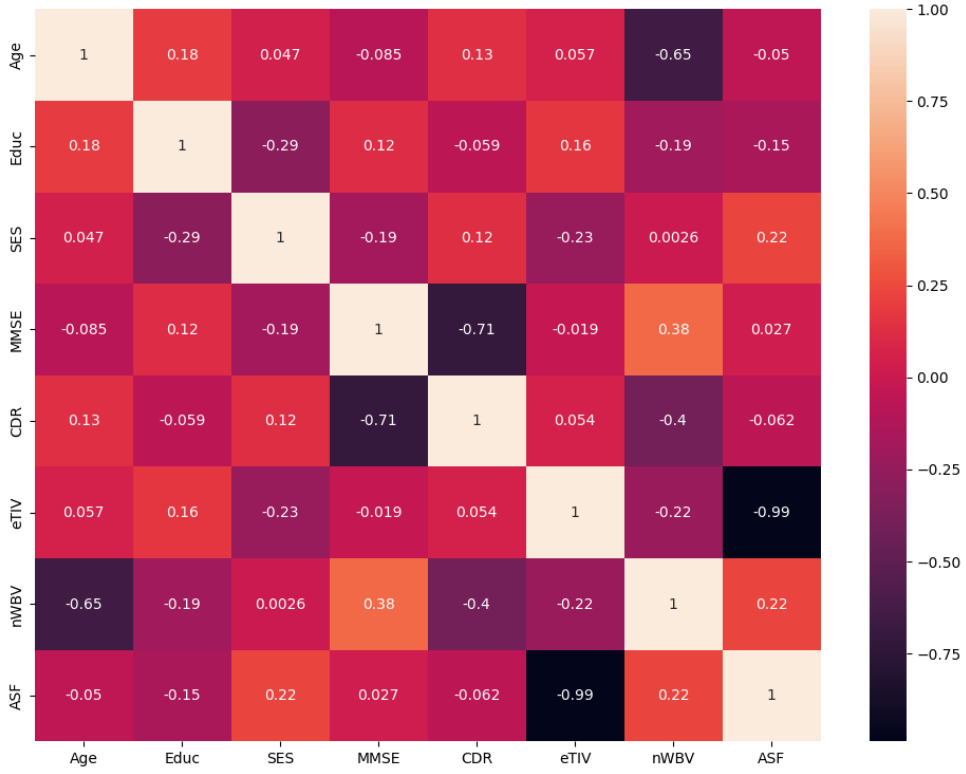


**Figure 8:** Heatmap describing correlations between features

Pairplots are a type of data visualization that displays scatter plots for pairs of continuous variables in a dataset. Here in Figure.9 we see the pair-plot figure which describes the relationship between all the continuous features in form of scatter plots. We color code all the data points using the target variable 'CDR'.

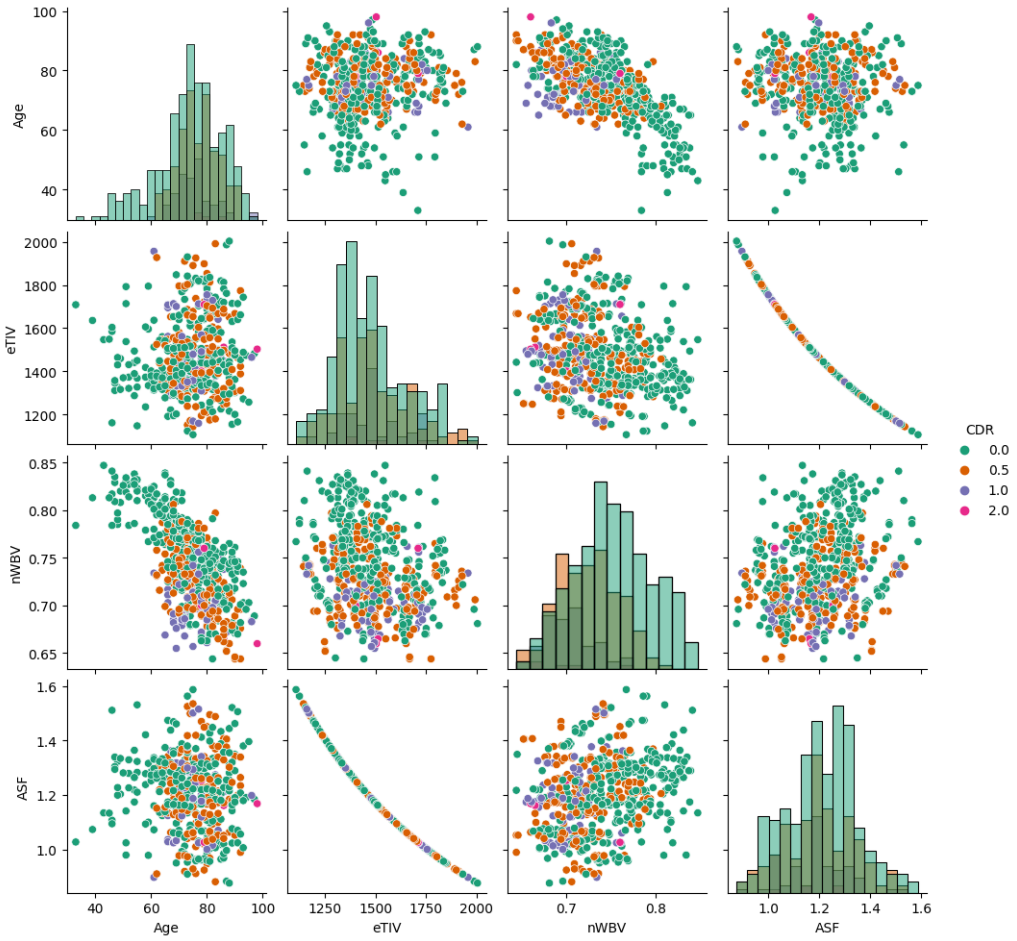


**Figure 9:** Pair-Plot describing the relation between each feature

### 3.4 Output Feature and its Interpretation

Clinical Dementia Rating (CDR): A rating scale that measures cognitive impairment severity, ranging from 0 (no dementia) to 3 (severe cognitive impairment).

- CDR 0: No dementia or normal cognitive function.
- CDR 0.5: Very mild stage of dementia, Early stages of cognitive impairment.
- CDR 1: Mild dementia, meaning mild cognitive impairment. May start affecting the functioning of the human body.
- CDR 2: Moderate dementia, indicating moderate cognitive impairment. Results in excess loss of functioning of human body
- CDR 3: Severe dementia, indicating severe cognitive impairment.

## 4. Implementation and Validation

### 4.1 ADS Systems

We are conducting the technical audit of two ADS systems as follows,

1. **Alzheimer's analysis using MRI:** [1]. The code for this ADS is taken from the submitted as a part of the Kaggle competition. This ADS utilizes both longitudinal and cross-sectional MRI data.
2. **Detecting Early Alzheimer's:** [2]. This code implementation is one of the best Kaggle competition submissions among the "MRI and Alzheimers" Kaggle series. It utilizes solely longitudinal MRI data.

The brief overview of the **Alzheimer's analysis using MRI** and **Detecting Early Alzheimer's** is as follows,

1. **Pre-processing Data:** After reading the longitudinal and cross-sectional MRI data, the columns are checked for the NULL values. The NULL values were only found in Education, SES, MMSE, and CDR. All the columns with NULL CDR values are dropped as CDR is our target variable. We also dropped redundant columns like ID and Delay.
2. **Data Concatenation:** The cross-sectional and longitudinal MRI data are concatenated from OASIS-1 and OASIS-2 respectively.
3. **Data Profiling:** The correlation heatmap is computed to better understand the dataset. Logistic Regression (w/ dropna)The graphs are drawn among input features to see the trends in correlation. Pearson's r, Spearman's $\rho$, Kendall's $\tau$, Phik $\phi\kappa$, and Cramer's V ($\phi c$) are utilized under correlation metrics.

1. Alzheimer's analysis using MRI
2. Detecting Early Alzheimer's

4. **Data Imputation:** To fill in the missing values in the "SES" columns with the most occurring data element. Similarly, we fill the missing values in the "MMSE" column with the median of that column.

5. **Data Normalization:** The CDR column is converted to categorical data using LabelEncoder and other numerical data is normalized.

6. **Test-train Split:** The concatenated dataset formed is then split in test and train split. The test split is having 30 percent of the total data.

7. **Model Training:**

   - **Alzheimer's analysis using MRI:** The hyperparameter tuning is done using cross-validation for XGBClassifier and GradientBoostingClassifier.
   - **Detecting Early Alzheimer's:** Logistic Regression (w/ imputation), Logistic Regression (w/ dropna), SVM, Decision Tree, Random Forest, and AdaBoost models were trained.

8. **Result Visualization:** The predictions are made using the best estimator while plotting the confusion matrix along with the classification report.

9. **Result Validation:** Accuracy, Recall, Precision, and Area Under the Curve metrics were recorded and checked to validate if the goals were met.

## 5. Outcomes

In the next part of our project, we will be working on auditing both our Automated Decision Systems described above using various fairness metrics. These metrics would be included

- Analyzing how our ADS is treating different gender groups. For example, is it treating both males and females equally or is it introducing a bias.

- We will be using LIME to analyze our ADS performance. LIME is very helpful in understanding how various features of a data set would influence the model's predictions for individual instances, providing insights into how and why our model is behaving in a specific manner.

- Next we can compute SHAP values for our Automated decision system. This would help us understand the extent to which different features influence the prediction and whether there are any unfair biases in the model's decision-making process due to these features. For example, in our dataset, we can find out the measure of the influence of gender, sex, etc on the target variable 'CDR'.

- We would also analyze the overall performance of all the classification models and why one performs better than the others. We can also check the fairness of both our ADS by evaluating the performance of the system across different subgroups or demographic groups.