Dementia Severity Prediction Across Multiple Levels using ML Algorithms

Presentation Overview

- Introduction
- Problem statement
- Business Understanding of the Predictive Model
- Dataset Explanation
- Dataset Preparation
- Exploratory Data Analysis
- Models of Decision
- Modeling Process
- Best Model Selection
- Future Scope

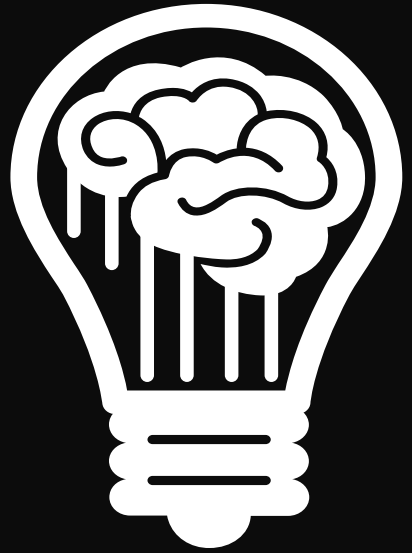
INTRODUCTION

Global Impact of Dementia

Dementia is a prevalent and impactful condition, affecting millions of individuals worldwide.

Recognizing Dementia

Common symptoms include memory loss, difficulty in communication, impaired judgment,



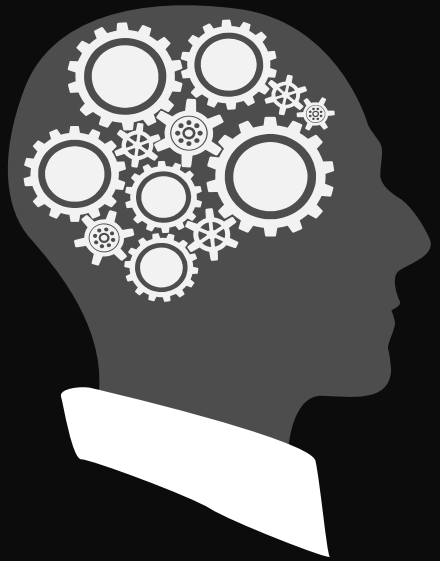
Diagnostic Complexity

Diagnosing dementia can be challenging, and there is currently no single definitive test.

Clinical Approach to Diagnosis

Healthcare professionals rely Melding Clinical, Medical, and Cognitive Assessments

Problem Statement

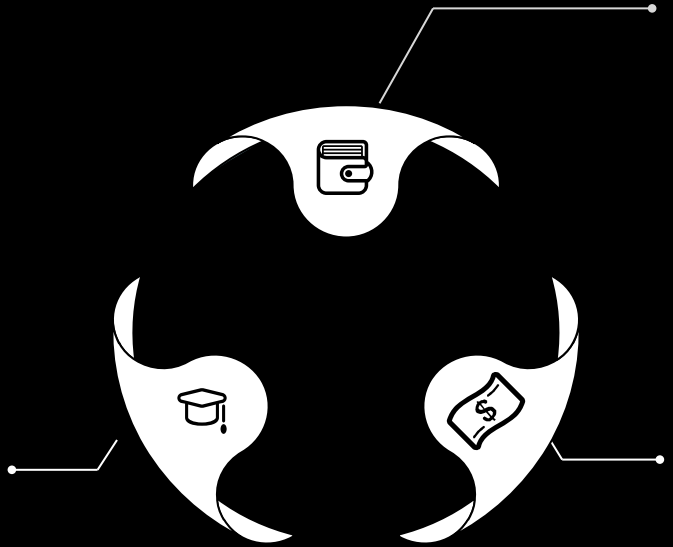


Develop a predictive model to estimate dementia severity using clinical data. Recognizing the significant number of undiagnosed cases, there's a critical need to utilize machine learning algorithms for early detection. Simplify complex profiles to facilitate early intervention and support in healthcare settings.

Business Understanding of the Predictive Model

Enhanced Diagnosis Precision

Reliable diagnostic tools reduce the likelihood of misdiagnoses, providing a more precise understanding of an individual's cognitive health status



Early Intervention & Treatment Planning

This contributes to improved patient outcomes and enhances the quality of dementia care.

Resource Optimization

Efficient allocation of resources allows healthcare providers to focus on specific interventions tailored to individual patient needs.

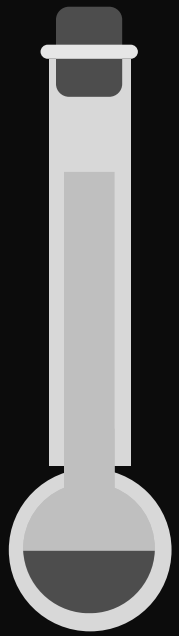
Dataset Overview

Oasis Cross-Sectional Studies dataset provided

by

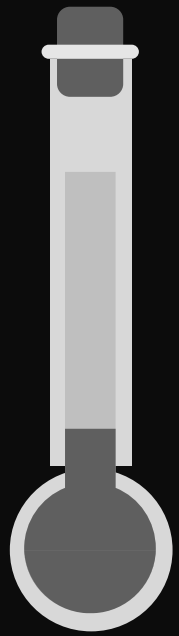
Open Access Series of Imaging Studies (OASIS)

Dataset Explanation



Identification and Demographics

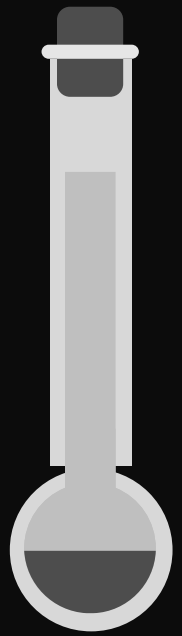
- **ID**: Unique identifier for each record or patient.
- **M/F:** Gender of the patients (Male or Female).
- **Age:** Age of the patients.



Educational and Socioeconomic Factors

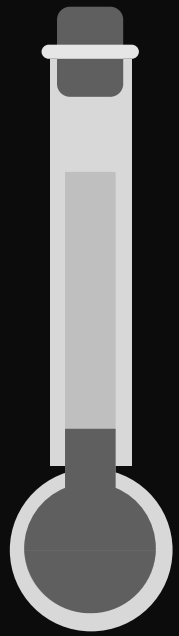
- **Educ**: Level of education of the patients.
- **SES**: Socioeconomic status of the patients.

Dataset Explanation



Cognitive Assessment Scores

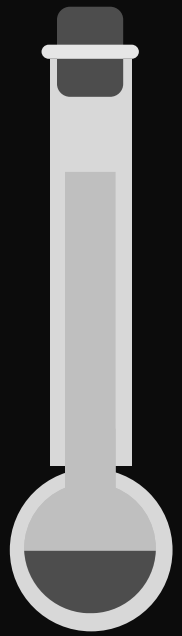
- **MMSE:** Mini Mental State Examination score is a brief and widely used cognitive screening tool to assess various aspects of cognitive function



Neuroimaging Measures

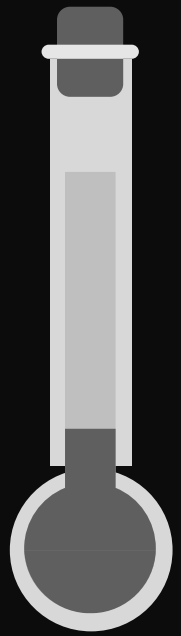
- **eTIV :** Estimated Total Intracranial Volume
- **nWBV :** Normalized Whole Brain Volume
- **ASF :** Atlas Scaling Factor

Dataset Explanation



Dominant Hand Information

- **Hand :** Represents the dominant hand of the patients (Right or Left)



Dementia Severity Target

- **CDR:** Clinical Dementia Rating, the target variable indicating the severity of dementia

Dataset Preparation

Data Collection

Gather diverse MRI scans and clinical data from sources like the **OASIS public archive** to cover varying levels of dementia severity.

Data Cleaning

Address **missing values and outliers** to ensure data integrity and prevent bias in the model

Data Transformation

Normalize/standardize numerical features for consistent model training.

Dataset Preparation

Data Splitting

Divide the dataset into **training, validation, and testing** sets for effective model evaluation.

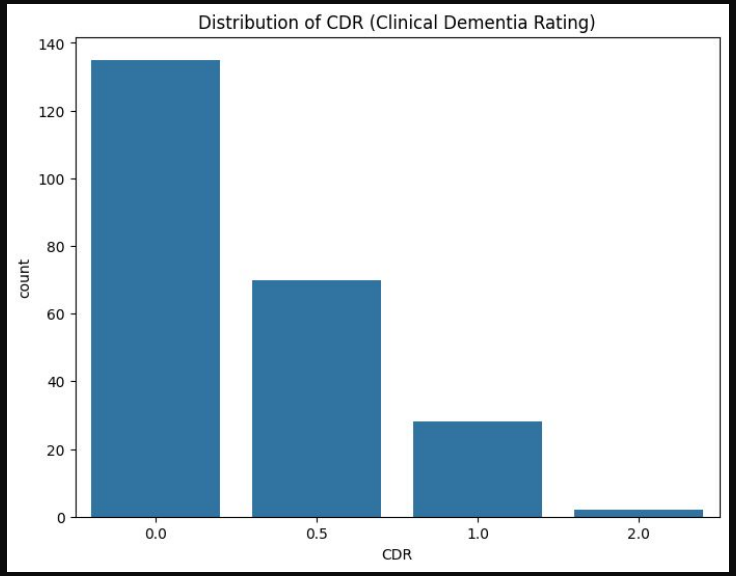
Data Encoding

Convert categorical variables using methods like **one-hot encoding** or **label encoding** for machine learning compatibility.

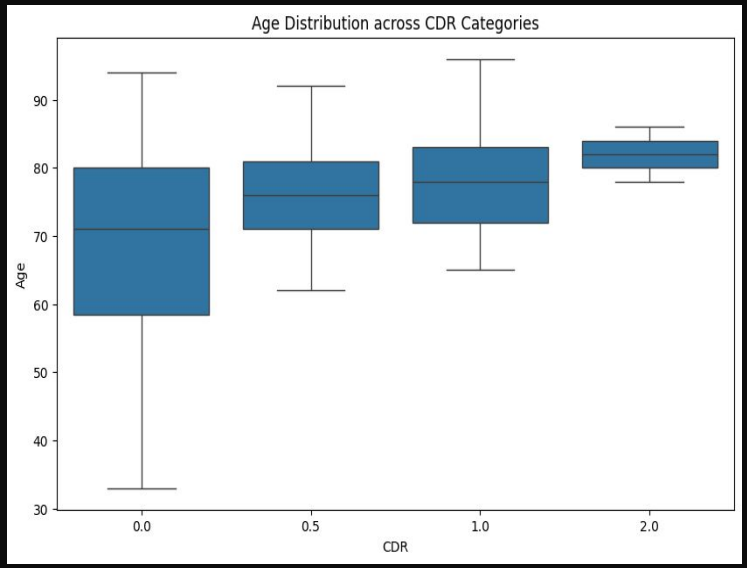
Handling Imbalanced Data

Mitigate imbalances in dementia severity distribution through techniques like **oversampling or undersampling.**

Exploratory Data Analysis

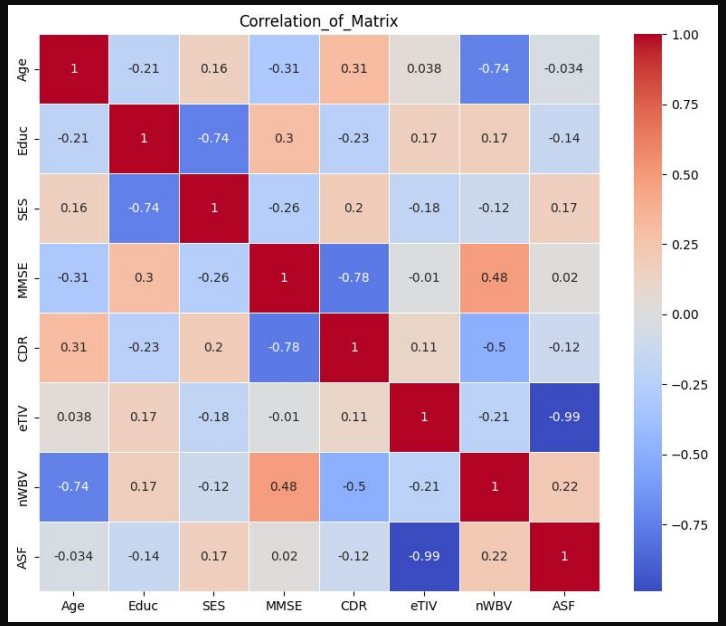


Histogram

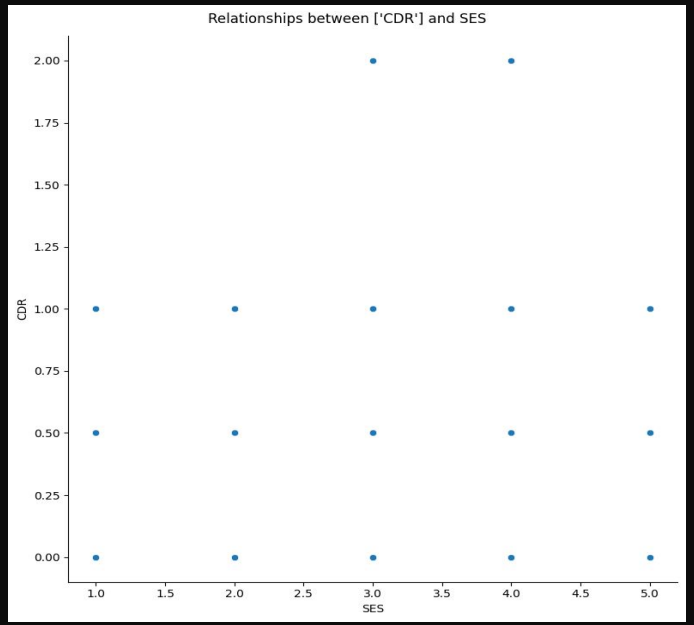


Box Plot

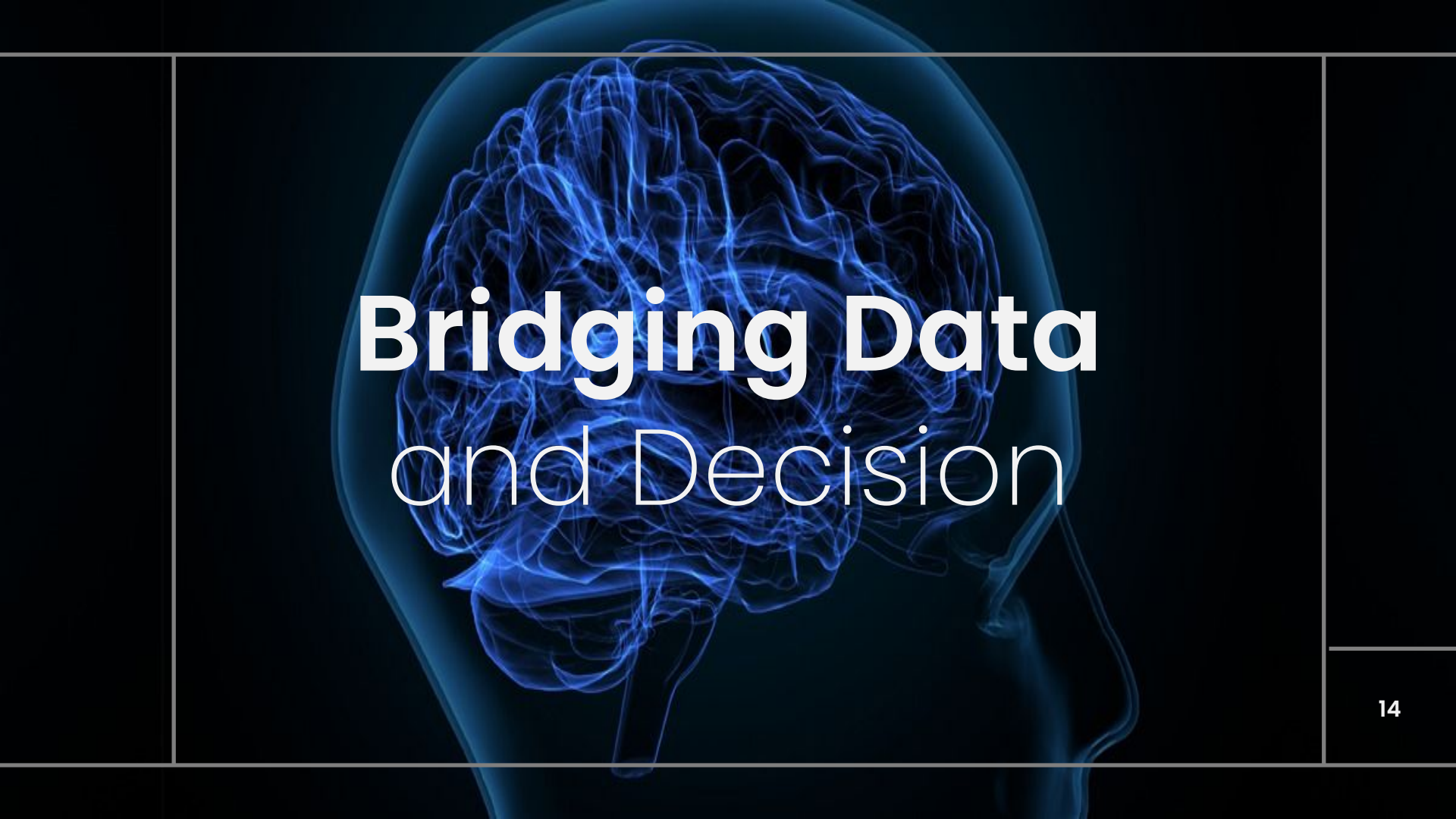
Exploratory Data Analysis



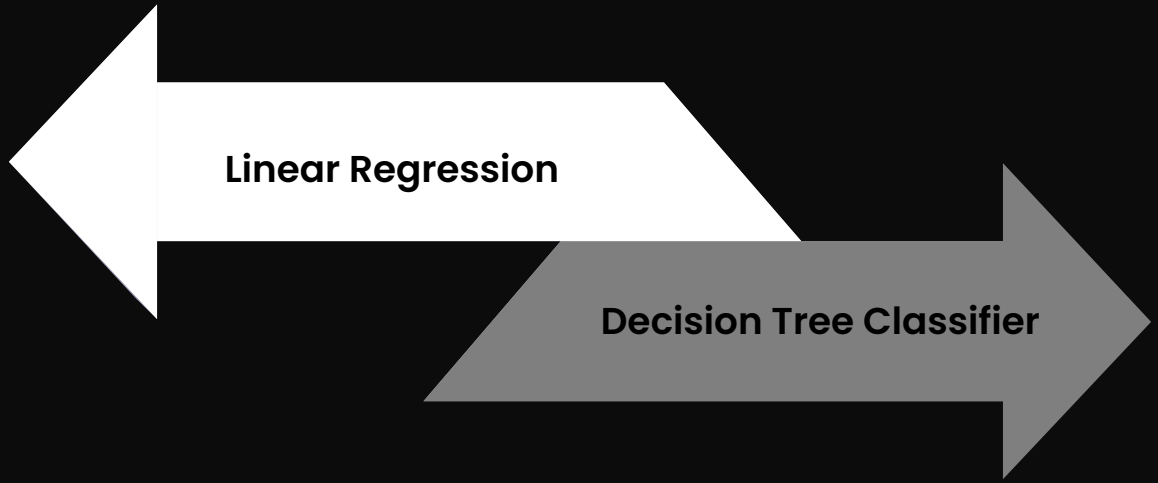
Correlation Matrix



Scatter Plot


Bridging Data
and Decision

Model’s of Decision



Modeling Process

Data Preprocessing

- Load data from a CSV file.
- Handle missing values by replacing them with the mean of the respective column.
- Encode categorical variables into numerical values.

Exploratory Data Analysis

- Identify dependent and independent variables.
- Analyze the correlation between different variables.
- Visualize positive and negative relationships using plots.

Data Scaling and Splitting

- Scale features using StandardScaler to standardize the dataset's features onto unit scale (mean = 0 and variance = 1).
- Split the data into a training set (80%) and a testing set (20%).

Modeling Process

Model Selection and Training

- Select models for the task: Linear Regression, Decision Tree Classifier
- Train each model using the fit() function on the scaled training data.

Model Evaluation

- For Linear Regression: Calculate Mean Squared Error (MSE).
- For Decision Tree Classifier: Calculate accuracy and generate a classification report.

Data Scaling and Splitting

- Evaluate the results of each model.
- Classify predicted values into various levels.
- Determine the best model based on evaluation metrics (MSE, accuracy, classification report).

Findings in Linear Regression

Mean Squared Error (MSE)

MSE of 0.0556 suggests a **precise fit, aligning model** predictions closely with actual values.

Coefficients Interpretation

'Age' coefficient at -0.000692 indicates a slight negative correlation with **Clinical Dementia Rating (CDR)**

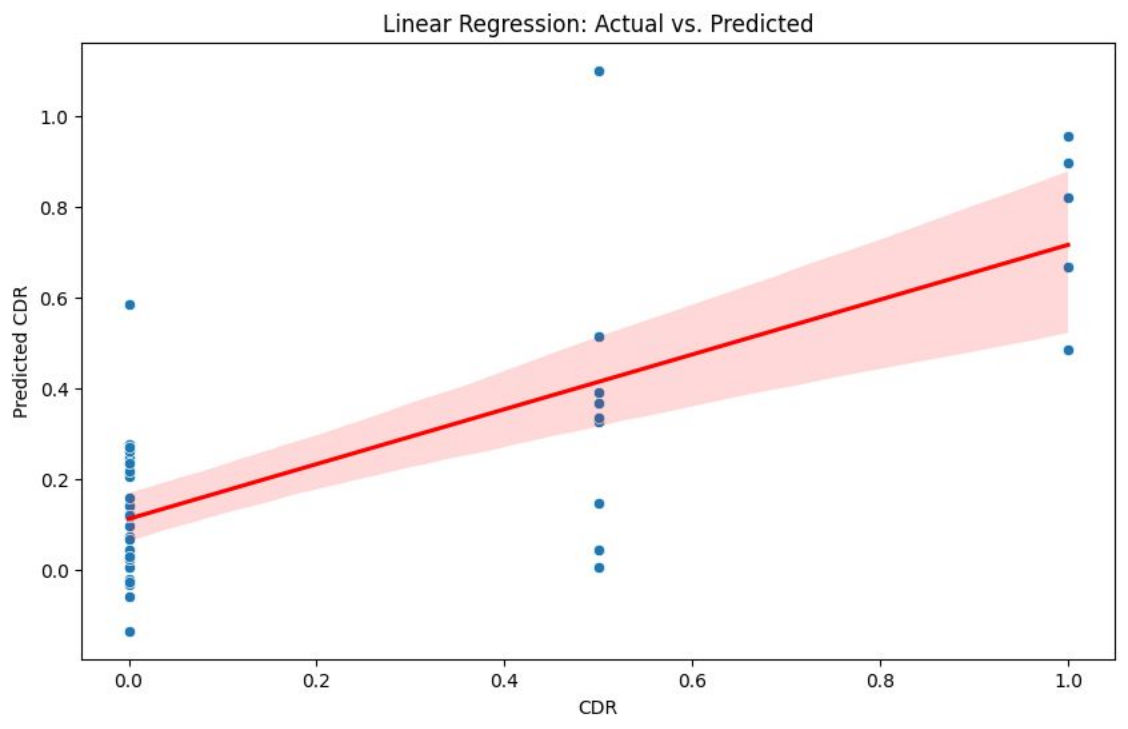
Intercept Value

Intercept at **5.1449 represents** the predicted **CDR when all features are zero.**

Protective Factors

Negative coefficients for **'Educ', 'SES', 'MMSE', 'eTIV', 'nWBV', 'ASF'** suggest **potential shields** against **dementia**.

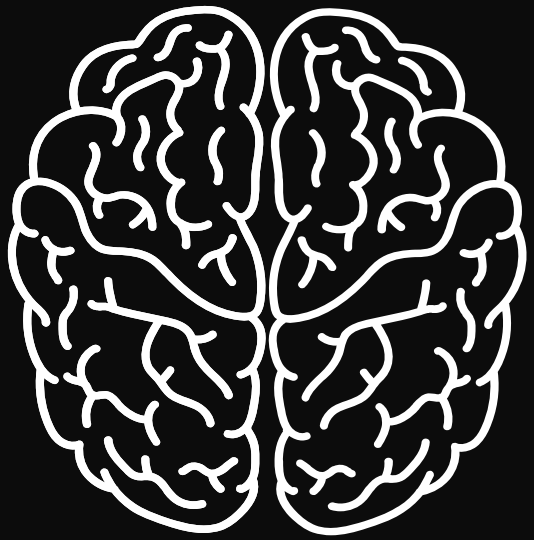
Findings in Linear Regression



Drawbacks in Linear Regression

Assumption of Linearity

Impact: Limits the model's ability to capture non-linear patterns in the data.



Sensitivity to Outliers

Impact: Outliers can disproportionately influence the model, affecting predictions.

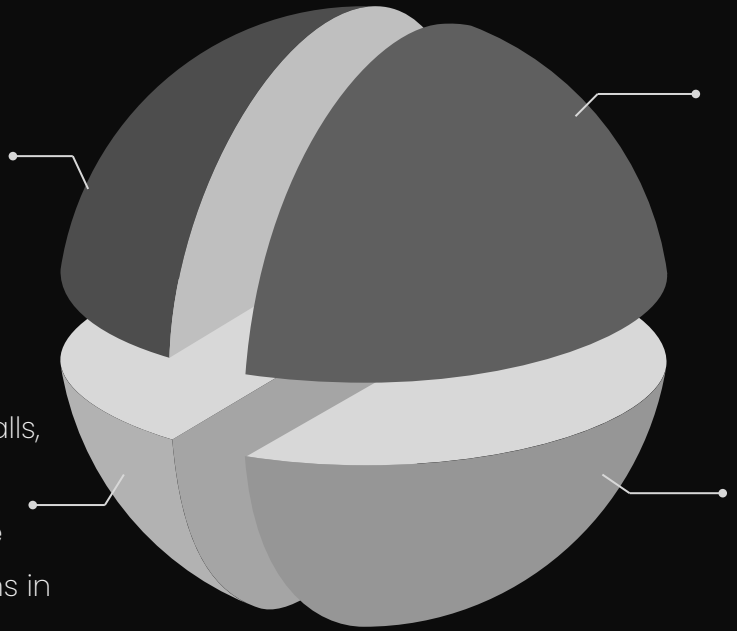
Findings in Decision Tree Classifier

Model Accuracy

77% accuracy in predicting dementia severity class showcases reliable performance

Recall Challenges

Challenges arise with low recalls, especially for "Mild Dementia" **(14%)** and none for "Moderate Dementia," revealing limitations in identifying these cases.



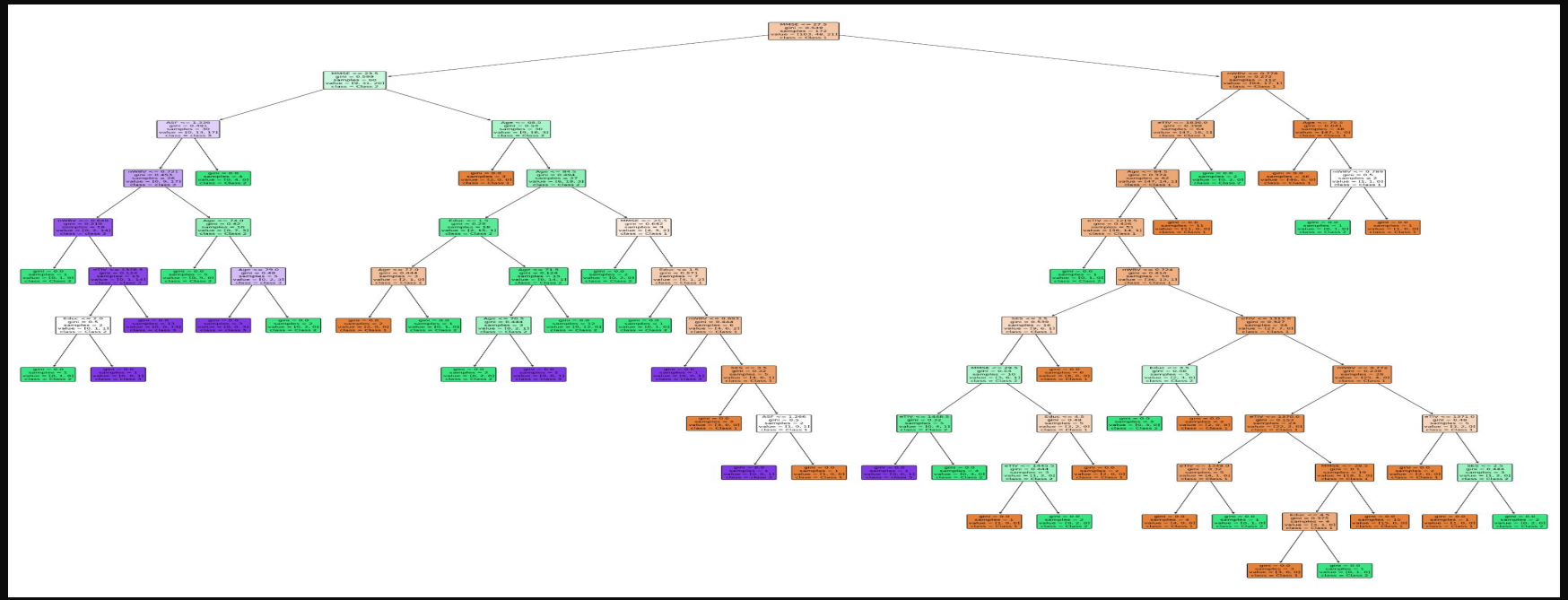
Precision and Recall

Class-specific precision **(e.g., 0.93 for CDR=0)** and recall **(e.g., 0.83 for CDR=0)** provide detailed insights into model efficacy.

Support Analysis

Examining the support column reveals the actual occurrences of each class in the test set, providing additional context for model performance

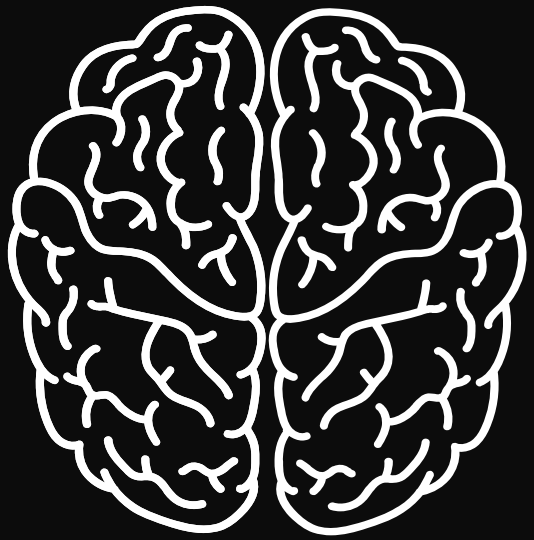
Findings in Decision Tree Classifier



Drawbacks in Decision Tree Classifier

Prone to Overfitting

Impact: Captures noise in the data, reducing the model's generalizability to new, unseen data.



Lack of Interpretability for Complex Trees

Impact: Difficulty in understanding and explaining intricate decision-making processes.

Conclusion of Findings Evaluation of models

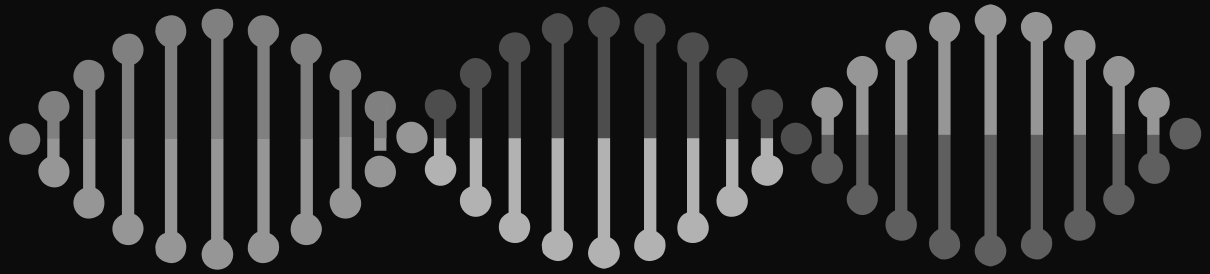
Linear Regression

A low MSE of 0**.0556** indicating a strong fit

Decision Tree Classifier

77% accuracy showcases reliable dementia severity class prediction performance.

Best Model Selection



Decision Tree Classifier

Given the likely non-linear relationships in healthcare data, the **Decision Tree Classifier** is more suitable for predicting dementia severity in our use case.

Future Scope

Collaboration with **healthcare professionals** to integrate advanced features positions the model for **real-world impact.**

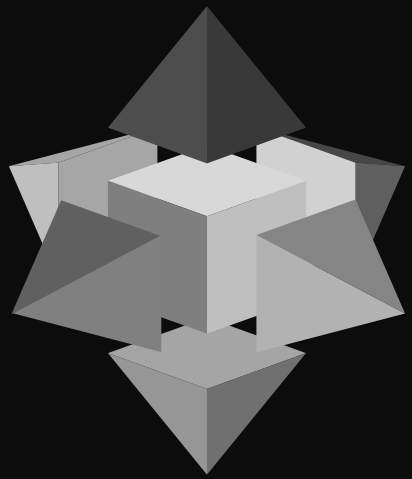
Incorporating **longitudinal data** contributes to **ongoing advancements** in personalized medicine.

THANK YOU

Any Questions?

TEAM MEMBERS AND ROLES

ADTA 5340
GROUP NO: 16



Vishnu Pasula

Data Visualization and exploratory data analysis

Chandana Vemula

Data preprocessing and Model evaluation

Vamshikrishna Thallapelli

Data Algorithms applications and prediction analysis