Prediction of Alzheimer

**Algorithms Used:**

* Random Forest
* Decision Tree
* K-Nearest Neighbors
* Logistic Regression
* XGBoost
* Catboost

**Colab Notebook:** <https://colab.research.google.com/drive/1D1GjPNPEeG5Mu8Elm_HxFrlpdbr1ocBZ?usp=sharing>

**Source:**

<https://www.kaggle.com/code/adhamtarek147/alzheimer-s-disease-prediction/notebook>

**Dataset:**

<https://www.kaggle.com/code/adhamtarek147/alzheimer-s-disease-prediction/input>

**Rows, columns:** (2149, 35)

**Initial Data Exploration Summary**

* The dataset contains **2,149 observations**.
* All values in the dataset are **non-null and numerical**.
* There are **no duplicate records**.
* After removing the DoctorInCharge and PatientID columns, the dataset consists of **33 features**.

**Observations from Visualization of Categorical Features**

* The dataset predominantly consists of individuals **without disease or health problems**.
* **Caucasian** is the most represented demographic.
* **High school graduates** constitute the largest educational group, followed closely by individuals with a **bachelor's degree**.
* Both **females and males** are equally represented across the dataset.

**Observations from Visualization of Numerical Features**

* Most of the columns show a **fairly uniform distribution**.
* The MMSE (Mini-Mental State Examination) scores exhibit a **bimodal distribution**, indicating two distinct groups within the data.
* The **heatmap** reveals no strong correlations among features.
* However, **five columns** show correlation with the target variable.

**Pearson Correlation Analysis**

* **Pearson correlation coefficient (Pearson's r):** Measures the linear relationship between variables, ranging from -1 to 1.
* **Significant correlations with the target variable:**
* **Negative correlations** (indicating lower scores are associated with a higher likelihood of Alzheimer's diagnosis):
* Functional Assessment: **-0.36**
* ADL (Activities of Daily Living): **-0.33**
* MMSE: **-0.24**
* **Positive correlations** (indicating presence of issues is associated with a higher likelihood of Alzheimer's diagnosis):
* Behavioral Problems: **0.22**
* Memory Complaints: **0.30**

This highlights the significance of these features in the diagnostic process.

**Algorithms Model and Accuracy**

|  |  |
| --- | --- |
| **Models** | **Scores** |
| Random\_forest | 0.93 |
| Decision\_Tree | 0.93 |
| K-Nearest Neighbors | 0.72 |
| Logistic Regression | 0.83 |
| XGBoost | 0.95 |
| CatBoost | 0.96 |

Best: **CatBoost**