Clemson University

Detection of Alzheimer's disease using Machine learning models.

checkpoint-2

CPSC 6300: Applied Data Science
Instructor: Dr. Nina Hubig
Semester: Spring 2023
Check Point - 1

BY

Jayanth Talasila
jtalasi@clemson.edu

Sreeram Paladugu
Spaladu@clemson.edu
C11281117

GitHub: https://github.com/sreerampaladugu10/ads-checkpoint-1

DecisionTreeClassifier model-

We selected the DecisionTreeClassifier model for predicting the diagnosis of individuals with Alzheimer's disease based on our exploratory data analysis (EDA). The DecisionTreeClassifier is a popular model for classification tasks due to its interpretability and effectiveness on diverse datasets.

During the EDA, we concluded that the key variables were Age, Sex, and Diagnosis, As a bar plot between Sex and Diagnosis suggested that sex may be a significant predictor of diagnosis. This insight provided valuable guidance for identifying crucial indicators for Alzheimer's disease, which informed our choice of the DecisionTreeClassifier model.

Model Evaluation-

The decision tree classifier was trained and evaluated on a dataset containing four different diagnosis categories: MCI, AD, HC, and AUD, with age and sex, also included as variables. The data was preprocessed by mapping the diagnosis categories to numerical values, splitting the subject-group column into MCI and AUD columns, and one-hot encoding the categorical features. The preprocessed data is then split into training and testing sets using a test size of 0.3 and a random state of 0.

After training the model, its performance was evaluated using accuracy, precision, recall, and F1 score. The model achieved an **accuracy of 0.7978** on the test set Its **precision score was 0.7902**. The **recall score was 0.7978**, and the **F1 score was 0.7710**.

The estimated test error rate of the decision tree classifier was obtained using 10-fold cross-validation. The best accuracy score obtained from it was used as an estimate of the test error rate. The estimated **test error rate was found to be 20.21%,** which indicates that the model fits the data relatively well, but it could be further improved.

```
Accuracy: 0.7978723404255319
Test error rate: 0.2021276595744681
Precision: 0.7902828230409963
Recall: 0.7978723404255319
F1 score: 0.7710642867751507
Best hyperparameters: {'max_depth': 10, 'min_samples_leaf': 1}
Best accuracy: 0.8766233766233767
```

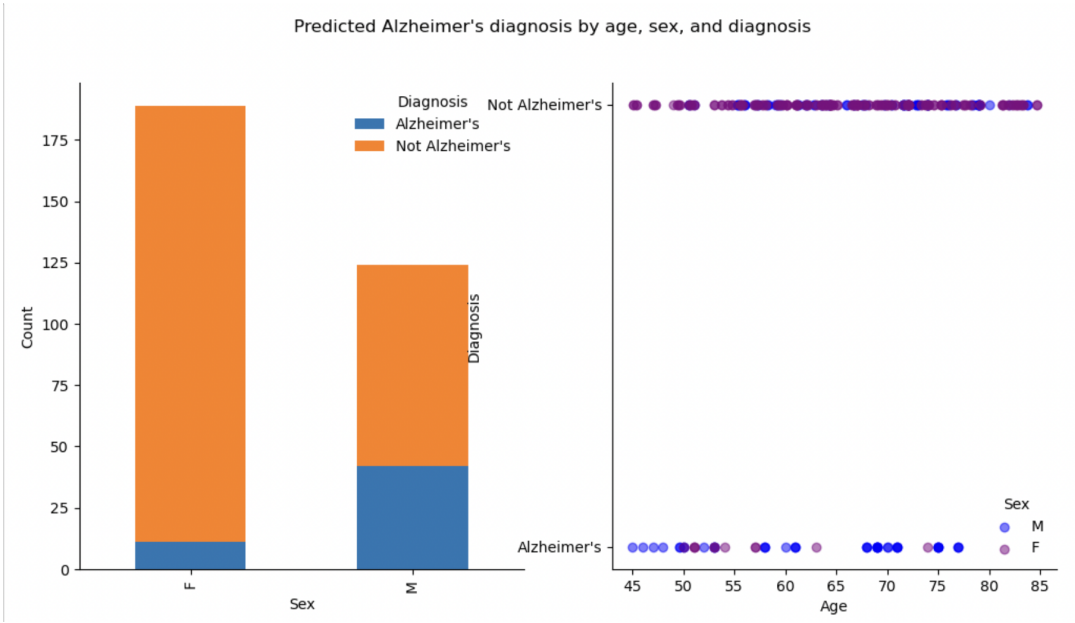
predictions and analysis-

We used our model to make a prediction on the new data point and print the predicted diagnosis (either Alzheimer's or Not Alzheimer's) for each combination of gender and age (for ages 45-85).

```
Predicted diagnosis for a M patient aged 45: Alzheimer's
Predicted diagnosis for a M patient aged 46: Alzheimer's
Predicted diagnosis for a M patient aged 47: Alzheimer's
Predicted diagnosis for a M patient aged 48: Alzheimer's
Predicted diagnosis for a M patient aged 49: Alzheimer's
Predicted diagnosis for a M patient aged 50: Alzheimer's
Predicted diagnosis for a M patient aged 51: Not Alzheimer's
Predicted diagnosis for a M patient aged 52: Alzheimer's
Predicted diagnosis for a M patient aged 53: Alzheimer's
Predicted diagnosis for a M patient aged 54: Alzheimer's
Predicted diagnosis for a M patient aged 55: Not Alzheimer's
Predicted diagnosis for a M patient aged 56: Not Alzheimer's
Predicted diagnosis for a M patient aged 57: Not Alzheimer's
Predicted diagnosis for a M patient aged 58: Alzheimer's
Predicted diagnosis for a M patient aged 59: Not Alzheimer's
Predicted diagnosis for a M patient aged 60: Alzheimer's
Predicted diagnosis for a M patient aged 61: Alzheimer's
Predicted diagnosis for a M patient aged 62: Not Alzheimer's
Predicted diagnosis for a M patient aged 63: Not Alzheimer's
Predicted diagnosis for a M patient aged 64: Not Alzheimer's
Predicted diagnosis for a M patient aged 65: Not Alzheimer's
Predicted diagnosis for a M patient aged 66: Not Alzheimer's
Predicted diagnosis for a M patient aged 67: Not Alzheimer's
Predicted diagnosis for a M patient aged 68: Alzheimer's
Predicted diagnosis for a M patient aged 69: Alzheimer's
Predicted diagnosis for a M patient aged 70: Alzheimer's
Predicted diagnosis for a M patient aged 71: Alzheimer's
Predicted diagnosis for a M patient aged 72: Not Alzheimer's
Predicted diagnosis for a M patient aged 73: Not Alzheimer's
Predicted diagnosis for a M patient aged 74: Not Alzheimer's
Predicted diagnosis for a M patient aged 75: Alzheimer's
Predicted diagnosis for a M patient aged 76: Not Alzheimer's
Predicted diagnosis for a M patient aged 77: Alzheimer's
Predicted diagnosis for a M patient aged 78: Not Alzheimer's
Predicted diagnosis for a M patient aged 79: Not Alzheimer's
Predicted diagnosis for a M patient aged 80: Not Alzheimer's
Predicted diagnosis for a M patient aged 81: Not Alzheimer's
Predicted diagnosis for a M patient aged 82: Not Alzheimer's
Predicted diagnosis for a M patient aged 83: Not Alzheimer's
Predicted diagnosis for a M patient aged 84: Not Alzheimer's
Predicted diagnosis for a M patient aged 85: Not Alzheimer's
```

```
Predicted diagnosis for a F patient aged 45: Not Alzheimer's
Predicted diagnosis for a F patient aged 46: Not Alzheimer's
Predicted diagnosis for a F patient aged 47: Not Alzheimer's
Predicted diagnosis for a F patient aged 48: Not Alzheimer's
Predicted diagnosis for a F patient aged 49: Not Alzheimer's
Predicted diagnosis for a F patient aged 50: Alzheimer's
Predicted diagnosis for a F patient aged 51: Alzheimer's
Predicted diagnosis for a F patient aged 52: Not Alzheimer's
Predicted diagnosis for a F patient aged 53: Alzheimer's
Predicted diagnosis for a F patient aged 54: Alzheimer's
Predicted diagnosis for a F patient aged 55: Not Alzheimer's
Predicted diagnosis for a F patient aged 56: Not Alzheimer's
Predicted diagnosis for a F patient aged 57: Alzheimer's
Predicted diagnosis for a F patient aged 58: Not Alzheimer's
Predicted diagnosis for a F patient aged 59: Not Alzheimer's
Predicted diagnosis for a F patient aged 60: Not Alzheimer's
Predicted diagnosis for a F patient aged 61: Not Alzheimer's
Predicted diagnosis for a F patient aged 62: Not Alzheimer's
Predicted diagnosis for a F patient aged 63: Alzheimer's
Predicted diagnosis for a F patient aged 64: Not Alzheimer's
Predicted diagnosis for a F patient aged 65: Not Alzheimer's
Predicted diagnosis for a F patient aged 66: Not Alzheimer's
Predicted diagnosis for a F patient aged 67: Not Alzheimer's
Predicted diagnosis for a F patient aged 68: Not Alzheimer's
Predicted diagnosis for a F patient aged 69: Not Alzheimer's
Predicted diagnosis for a F patient aged 70: Not Alzheimer's
Predicted diagnosis for a F patient aged 71: Not Alzheimer's
Predicted diagnosis for a F patient aged 72: Not Alzheimer's
Predicted diagnosis for a F patient aged 73: Not Alzheimer's
Predicted diagnosis for a F patient aged 74: Alzheimer's
Predicted diagnosis for a F patient aged 75: Not Alzheimer's
Predicted diagnosis for a F patient aged 76: Not Alzheimer's
Predicted diagnosis for a F patient aged 77: Not Alzheimer's
Predicted diagnosis for a F patient aged 78: Not Alzheimer's
Predicted diagnosis for a F patient aged 79: Not Alzheimer's
Predicted diagnosis for a F patient aged 80: Not Alzheimer's
Predicted diagnosis for a F patient aged 81: Not Alzheimer's
Predicted diagnosis for a F patient aged 82: Not Alzheimer's
Predicted diagnosis for a F patient aged 83: Not Alzheimer's
Predicted diagnosis for a F patient aged 84: Not Alzheimer's
Predicted diagnosis for a F patient aged 85: Not Alzheimer's
```

We also plotted the prediction of Alzheimer's based on age and sex where we can clearly see that the likelihood of men being predicted to have Alzheimer's is higher compared to women.



conclusion -

Overall, it appears that using the important factors of age, sex, and diagnosis, the DecisionTreeClassifier model performs reasonably well in predicting the diagnosis of Alzheimer's disease. The algorithm was able to estimate the chance of Alzheimer's disease for new data points based on age and sex and received an accuracy score of 0.7978 on the test set.

It's intriguing to note that the model projected that male patients would have a higher likelihood of having Alzheimer's disease than female patients, It's crucial to remember that this is only a forecast and not a confirmed diagnosis and that further testing and analysis may be required to confirm or rule out Alzheimer's disease.

Although the model fits the data reasonably well, with an estimated test error rate of 20.21%, there is still space for improvement. To increase the model's precision and predictive capability.

Overall, the DecisionTreeClassifier model offers a helpful method for estimating the chance of Alzheimer's disease based on age, sex, and diagnosis and may be applied as a component of a more comprehensive procedure for the disease's screening and diagnosis.