Alzheimer’s Disease Prediction

TESSA CLARY

Questions Sought to Answer

- This project set out to explore prediction factors of Alzheimer's Disease.
  - How do various demographic factors influence cognitive decline?
  - Which variables among age, gender, and race show a significant impact on cognitive decline?

Data Preparation Work

- Pandas was used to find and drop rows with missing values.
- Column attributes were analyzed and those found to be redundant or unnecessary were removed from the dataset.
- The original dataset was split into three derivative datasets:
  - Age
  - Gender
  - Race
- Dummy variables were created in order to encode the categorical variables provided into Boolean values that could be used by the models.

Tools Used

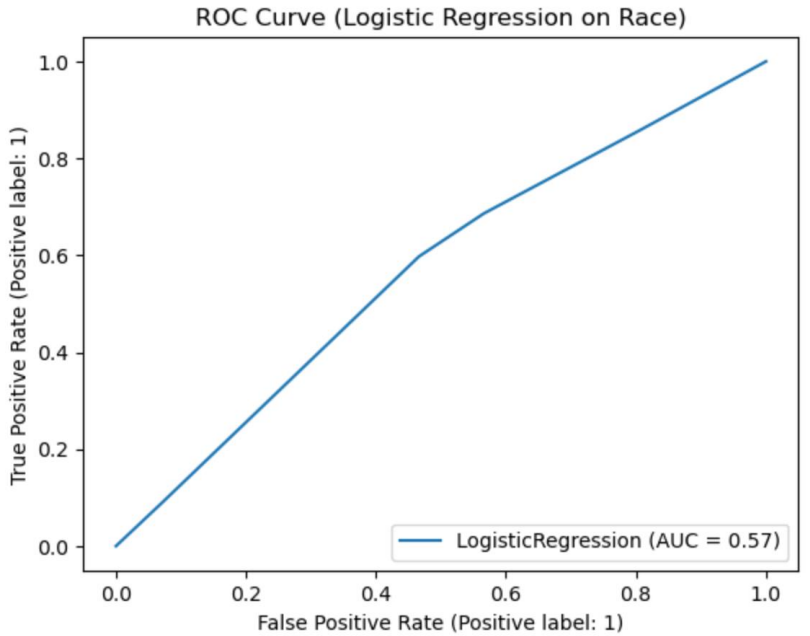
- This project relied on Pandas for dataset usage and manipulation.
- Several different Python machine learning libraries were used in an attempt to glean as much data as possible from the datasets:
  - Logistic Regression
  - Decision Tree Classifiers
  - Gaussian Naïve Bayes
- These alongside metrics for accuracy, precision/recall, and ROC curves were used from sklearn.

Classification Techniques Applied

- Logistic Regression
  - This model attempts to fit the encoded dummy variables into a logistic regression model. The coefficients provided from this model help determine the impacts of the given independent variables on the dependent variable of cognitive decline.

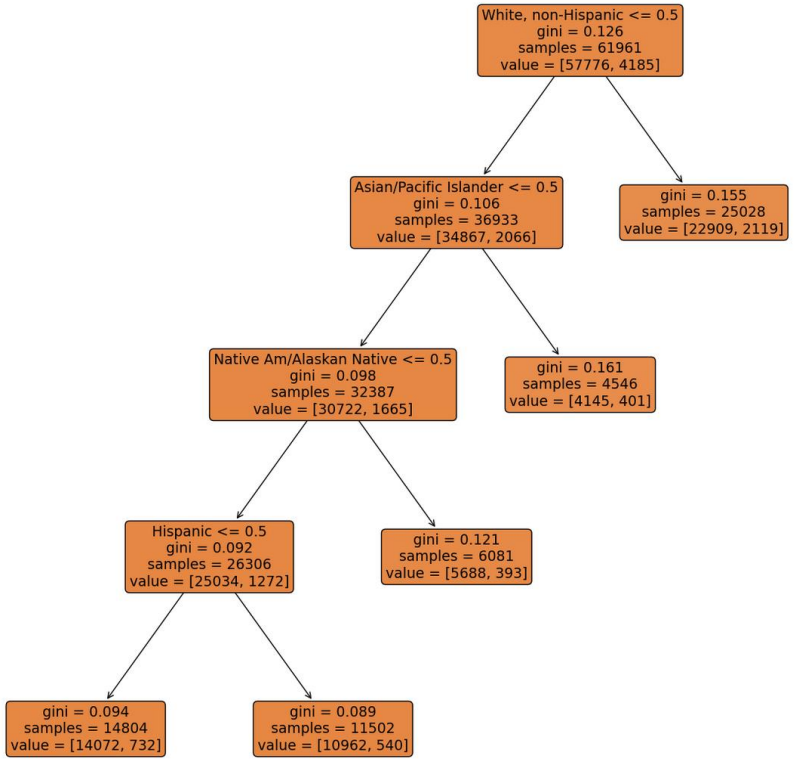
Accuracy: 0.9351881737783229

	Feature	Coefficient
0	Asian/Pacific Islander	0.334463
1	Native Am/Alaskan Native	-0.001227
2	Black, non-Hispanic	-0.284679
3	Hispanic	-0.338887
4	White, non-Hispanic	0.290339



Classification Techniques Applied (cont.)

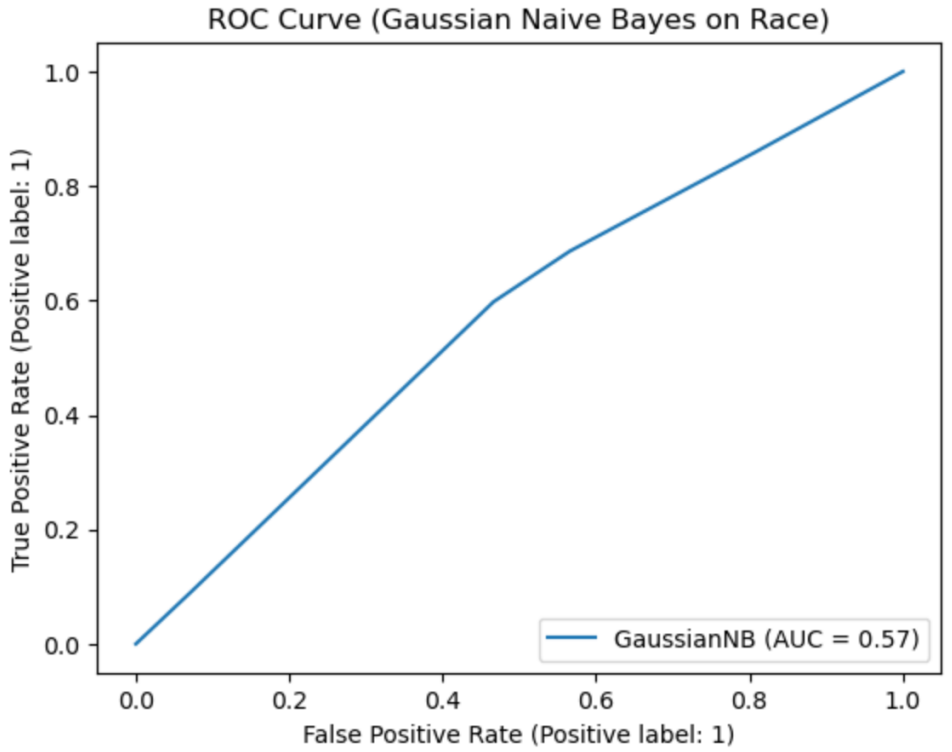
- Decision Tree Classifier
  - This model used the dummy variables to display a tree-like structure that modeled the relationships and patterns between the independent variables and cognitive decline. Its visualization helps aid understanding as to how the different variables contribute to predicting cognitive decline.



Classification Techniques Applied (cont.)

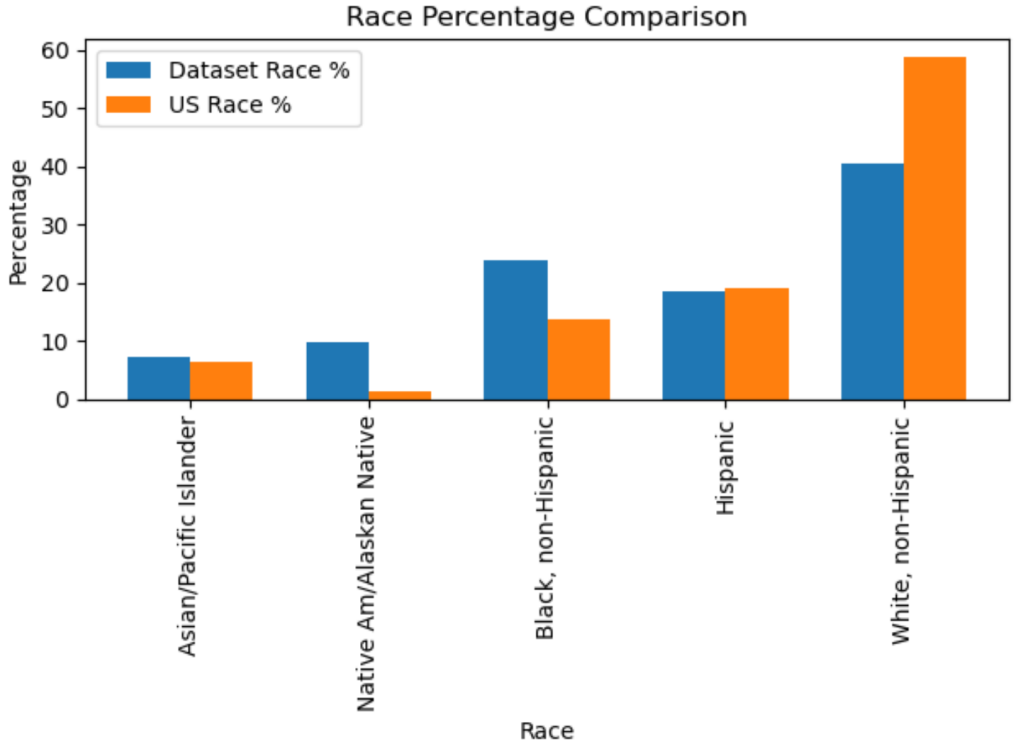
- Gaussian Naïve Bayes
  - This model uses Bayes theorem with the assumption of independence between features. In this study, it calculated the probability of cognitive decline for each independent variable.

Accuracy: 0.9351881737783229



Knowledge Gained

- Race was found to be a notable influential factor in predicting cognitive decline as it pertains to Alzheimer's Disease.
- Age and gender were found to have minimal impact on the dependent variable of cognitive decline in comparison.
- Further questions regarding BRFSS data collection methodology have been raised due to race data not necessarily reflecting the U.S. population accurately.



Application of Knowledge

- This study found that race may play an influential role in predicting cognitive decline due to Alzheimer's Disease.
  - This is hampered by the fact that other well known influential factors such as age and gender did not show any predictive capability in comparison.
  - On the basis of race impacting cognitive decline, this can have a number of possible impacts:
    - The ability to diagnose may improved with the additional consideration of race as a demographic variable.
    - At risk individuals may be easier to identify in the earlier stages of progression.
    - Other research may better be able to focus their resources on these known at risk populations.\
- The validity of the dataset accurately representing the population it is meant to study is called into question; the CDC should consider updating their data collection methodology to better reflect the actual U.S. populations.