**Dataset used for analysis:** Oasis Longitudinal Dataset

This set consists of a longitudinal collection of 150 subjects aged 60 to 96. Each subject was scanned on two or more visits, separated by at least one year for a total of 373 imaging sessions. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included. The subjects are all right-handed and include both men and women. 72 of the subjects were characterized as nondemented throughout the study. 64 of the included subjects were characterized as demented at the time of their initial visits and remained so for subsequent scans, including 51 individuals with mild to moderate Alzheimer’s disease. Another 14 subjects were characterized as nondemented at the time of their initial visit and were subsequently characterized as demented at a later visit.

[Source: <https://www.oasis-brains.org/>]

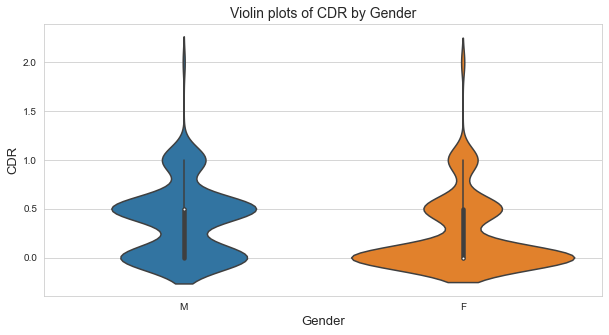
The features in the analysis are:

1. MR Delay: Delay between the first and second MRI Scans.
2. Gender: Male/Female
3. Orientation of hand movement basis: Right Handed/Left handed
4. Age
5. EDUC: Years of Education
6. SES: Socioeconomic Status (Socioeconomic status as assessed by the Hollingshead Index of Social Position and classified into categories from 1 (highest status) to 5 (lowest status))
7. MMSE: Mini-Mental State Examination score (range is from 0 [worst] to 30 [best]) (Folstein, Folstein, & McHugh, 1975)
8. CDR: Clinical Dementia Rating. (0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2. = moderate AD)
9. ASF: Atlas scaling factor (unitless). Computed scaling factor that transforms native-space brain and skull to the atlas target (i.e. the determinant of the transform matrix)
10. eTIV: Estimated total intracranial volume
11. nWBV: Normalized whole brain volume, expressed as a percent of all voxels in the atlas-masked image that are labeled as gray or white matter by the automated tissue segmentation process

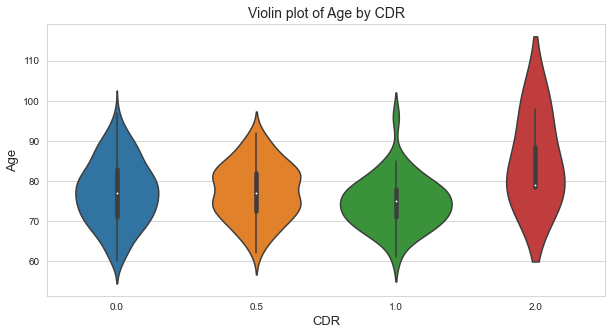
**EDA:**

From the analysis of CDR, we see that:

1. Dementia is more common in Men,



1. Dementia is more common in people of age range 70 to 80,



**ML Prediction algo results:**

Data preprocessing was done using Autoencoders and Labelencoders, prewritten in SKLEARN.

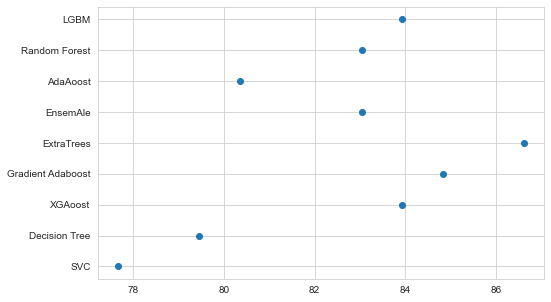
We used various models for prediction.

We used:

* GAUSSIAN CLASSIFIER
* SVC
* Decision Tree
* XGBoost
* Gradient Boost
* Bagging
* Adaboost
* Random Forest
* LGBM
* Naive Bayes

The accuracy of prediction of Dementia is put in the following table:

|  |  |
| --- | --- |
| Method | Accuracy |
| SVC | 77.67 |
| Decision Tree | 79.46 |
| XGBoost | 83.92 |
| Gradient Boost | 84.82 |
| ExtraTrees | 86.61 |
| Ensemble | 83.05 |
| Adaboost | 80.35 |
| Random Forest | 83.05 |
| LGBM | 83.92 |

**ROC Curve of analysis is below:**

**(**[**https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc#:~:text=An%20ROC%20curve%20(receiver%20operating,False%20Positive%20Rate)**](https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc#:~:text=An%20ROC%20curve%20(receiver%20operating,False%20Positive%20Rate))

An **ROC curve** (**receiver operating characteristic curve**) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

* True Positive Rate
* False Positive Rate

