This homework aims to review and practice fundamental machine learning concepts. The idea is to build a predictive model of whether a respondent likely voted in their last presidential election. For this purpose, the "cses4\_cut.csv" file is used containing a subset of the CSES Wave Four data set.

As requested, different models and approaches have been tested. Here is a shortlist of what has been implemented:

Text

Description automatically generated

**Dimensionality-reduction with feature selection**

We can reduce overfitting, improve accuracy, and reduce training time with feature selection, for this purpose, “sklearn.feature\_selection.SelectKBest” is used, and 12 features were with the highest score.

Pre-processing:

There are some unwanted data in the data set like:

1. REFUSED
2. VOLUNTEERED: DON'T KNOW
3. MISSING

A picture containing text, monitor, wall, indoor

Description automatically generated

These data disrupt the distribution of data. I used the quantile transformer method “sklearn.preprocessing.QuantileTransformer” to solve this problem. This method transforms the features to follow a uniform or a normal distribution. Therefore, this transformation tends to spread out the most frequent values for a given feature. It also reduces the impact of outliers.

A picture containing graphical user interface

Description automatically generated

Classifiers with dimensionality-reduction and preprocessing

After preprocessing and feature selection, I re-trained the models. Results are as follows:

Text

Description automatically generated

Optimizing the model and its hyperparameters

I took the top 5 classifiers and regressors and looped them until I found the best hyperparameters.

Results are as follows:

Text

Description automatically generated