In this project, I built a predictive model to understand the likelihood of a respondent to vote in the last presidential election.

I used 'cses4cut.csv' file which includes a subset of the CSES Wave Four data set. Different classifiers and regressors were tested in this project.

I have tried GaussianNB, Logistic Regression, KNeighbors, Random Forest and SVC to find the best classifier fit for my data. One of the best performing algorithms, according to the classification reports, was the Random Forest with a classification report.

Table 1 demonstrates that the findings are in line with the result.

Figure 1: The following results were yielded.

	Model	Accuracy
4	Random Forest	86.69%
1	K-Nearest Neighbors	84.47%
2	Linear Discriminant Analysis	83.75%
0	Logistic Regression	83.27%
3	Quadratic Discriminant Analysis	69.86%
5	Bayes	69.34%

1 ML Algorithm

Initially, I've tried the following four algorithms from Scikit-Learn: GaussianNB Classifier, Random Forest Classifier, KNeighborsClassifier.

In all the algorithms I used in the project, people who say they did not vote in the elections are not classified. I guess it may be my failure. However, in the dataset, voting and non-voting people can be estimated in the voting results. (True:10226, False:2225)

Figure 2: GaussianNB Classifie.

	precision	recall	f1-score	support
False	0.30	0.55	0.39	719
True	0.88	0.73	0.80	3390
accuracy			0.70	4109
macro avg	0.59	0.64	0.59	4109
weighted avg	0.78	0.70	0.73	4109

There does not seem to be much difference in the scores obtained in the GaussianNB Classifier, Random Forest Classifier and KNeighbors Classifier algorithms.

Figure 3: Random Forest Classifier.

	precision	recall	f1-score	support
False	0.63	0.23	0.34	719
True	0.86	0.97	0.91	3390
accuracy			0.84	4109
macro avg	0.75	0.60	0.63	4109
weighted avg	0.82	0.84	0.81	4109

All models except the KNeighbors Classifier resulted in overfitting.

Figure 4: KNeighbors Classifier.

	precision	recall	f1-score	support
False	0.52	0.02	0.04	719
True	0.83	1.00	0.90	3390
accuracy			0.83	4109
macro avg	0.67	0.51	0.47	4109
weighted avg	0.77	0.83	0.75	4109