

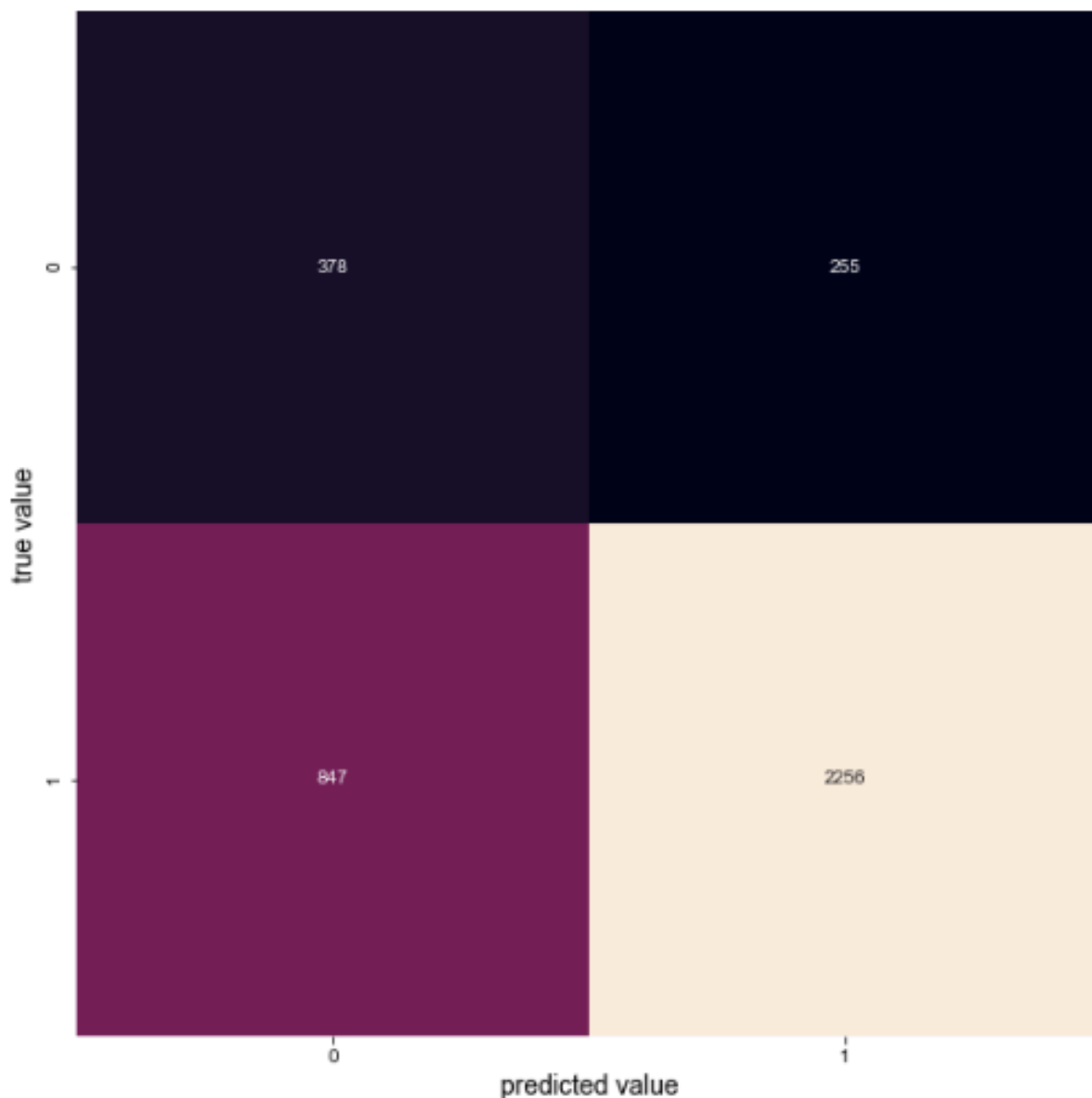
### Homework 3

In this project, the aim is to predict who will vote or not vote depending on some demographic information. The steps that are taken stated as below:

1- After reading the data, I used one hot encoding process to convert categorical variables into a form to be used in prediction.

2- I splitted the data into two part: %70 train and %30 test

3- Using Gaussian Naive Bayes, I applied bayes theorem to predict the outcome. The accuracy for this prediction is 0.70. The confusion matrix is shown as below.



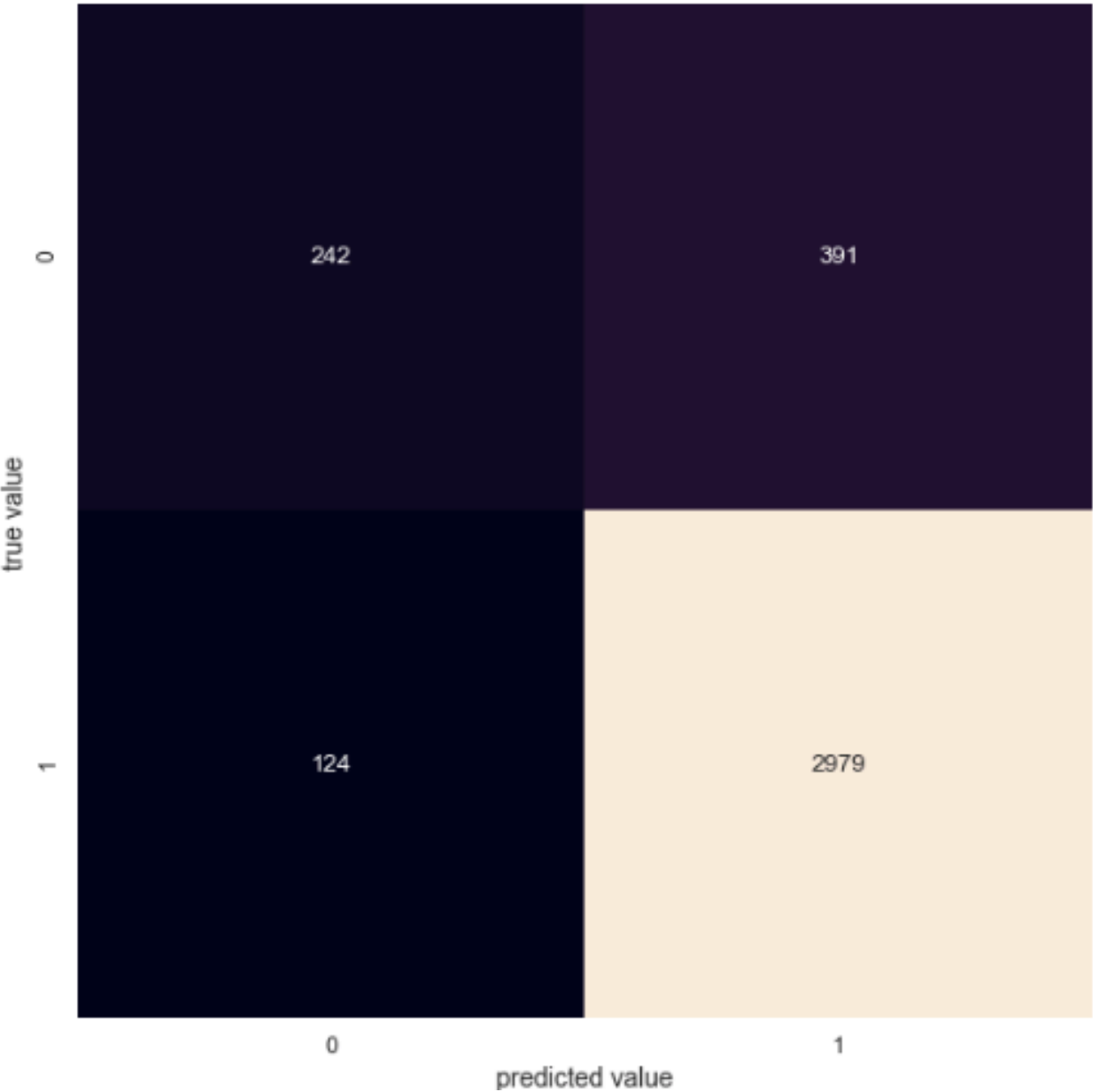
4- To

increase model efficiency, I made feature selection with k highest score. Then, I chose most effective parameters on the model and made a prediction again.

'D2026', 'D2027', 'D2029', 'D2021', 'D2011', 'D2030', 'D2022', 'D2028',  
'D2023', 'D2015', 'age', 'D2016']

With this improvement, model is saved from overfitting and resulting accuracy is 0.74

5- Because feature selection gives an increase in the accuracy of the model, I would like to try to model with another algorithms. With decision tree and random forest, I made predictions on the data. The resulting accuracies are 0.78 and 0.86 respectively. The confusion matrix after random forest is shown below and it is clear to see the improvement in true prediction.



Hande Kastan  
0065740