HomeWork3 Report

Fort he third homework I tried to predict whther or not a person voted via 3 explanotary variables that are educational level, socioeconomic status and househlod income. Fort hat I have imported the boolean y values first and converted them to a numpy array. After that I imported my explanotary variables and concanated them together. After I got rid-off the nan values that I set fort he uncertain answers, I scaled my data for a better process. I have implemented 3 different ML models. First one is Logistic Regression which we didnot cover in the lectures. I've heard of it during a conversation with a friend of mine, who is a data scientist. I was told that Logistic Regression is one of the best fits fort he task I have. My accuracy score was 83%. After that I implemented Gaussian Naive Bayes due to the simplicity of my variables. I got the exact same result with the previous model and I have implemented a 10-fold cross validation. The scores was very close to each other so that I understand that the model offers stabilized outputs. My third model was the GMM. Here, I reached a 82% accuracy score, however, as I run the code over and over again, I discovered that the accuracy score declined around 2% at the end. I believe that this result is due to the nature of the model unless it is a consequence of the usage of same kernel repeatedly. When I restart the kernel, the model kept giving relatively higher accuracy scores. Since I tried three different models and the maximum accuracy scores are very close to each other, I assume that my choice of explanotary variables could have been better. The variables predict the target at a 83% accuracy rate. It could have been over 90% if I would try with other variables. In the classes I have seen the models that predict over 96% accuracy rate. Although my accuracy rate is not very high, the process and the hands-on experience of machine learning was generally successful in my opinion. I had the opportunity to try different models.