Q2: Import HistoricMortages.csv

This dataset is from Freddie Mac and includes characteristics of mortages that Freddie Mac has purchased in the past from banks. The dataset also includes a binary variable indicating if each mortage entered default at any time during the lifetime of the loan. Freddie Mac would like you to develop a machine learning algorithm to predict the probability of default for future loans. They would like an algorithm or algorithms that will allow them to make decisions about what mortages to buy in the future.

1. [10 Points] Develop a model that generates a classification prediction. That is, a model that will assign each loan to either the category of default or non default. Provide evidence for the performance of you model and make an argument for why you think your model is the best fit model that you could develop given what we have learned so far. Support your argument with any graphs or information that you may need. Save your final model for deployment.

Import NewMortages.csv

In the NewMortages.csv dataset, you will find unlabeled data, that is, you will find variables that define characteristics of new mortages that Freddie Mac is considering purchasing, but, because these are new loans, they do not have data on default. Using your models, answer the following questions to help Freddie Mac know what loans to purchase and which loans to not purchase.

2. [5 Points] Generate a predicted classification for each mortage in the new dataset. What do you predict will be the overall default rate if you buy all the mortages (Report the mean of the predicted default probability for this)?

3. [5 Points] Say you want to approve as many loans as possible but maintain an over all default rate of no greater than 7%. Generate a variable that labels what mortages will be purchased and which ones will not. (You may need to create a predicted probability of default and use some threshold of predicted probability to answer this question.) How many loans would you purchase and how many would you reject?

4. [10 Points] Finally, optimize the the acceptance decision so that you maximize profits, instead of trying to set the overall default rate.

1. Calculate the expected profit for each loan. "Expected profit" means the probability that the loan is repaid times the profit earned if repaid plus the probability that the loan is not repaid times the profit if the loan is not repaid. Assume that the loan is repaid using a standard amortized loan repayment structure with the principle provided in the Original.UPB, using the monthly interest rate provided in the interest rate variable, for a duration of the loan (in months) provided in the duration variable. Assume that the bank must buy the loan for the cost of the principle (so they incure that cost regardless of whether the loan is repaid). Also assume that if someone defaults that they default entirely so nothing is repaid in that case (I admit that this is an extreme assumption).

2. When you are maximizing profits, What will your profits be, how many loans will you approve, and what is the loan repayment probability cutoff that you used to optimize profits?

3. If you were a policy maker who wanted to implement a policy that keeps banks from assuming too much risk, porfolio risk in excess of 7%, would your need to impose a portfolio risk limit in this case? That is, would you need to 'regulate' Feddie Mac to keep them from assuming too much risk?

4. Let's say that we want to increase access to affordable housing. We could increase access to housing by subsidizing the mortage market. If Freddie Mac can buy mortages at a reduced price, then perhaps they could buy more mortages which would in term create an incentive for banks to offer more mortages; this would make it easier for people to buy homes. So let's subsidize the cost of purchasing a mortage by saying that the government will cover 40% of the cost of the mortages that Freddie Mac purchases. Maximize the profits again. What are the profits, number of loans purchased, and the default rate at the new subsidized optimal?

5. If we are subsidizing the mortage market as described in 4.D, would it now become necessary to regulate the amount of risk that Freddie Mac takes on in order to keep them from holding a porfolio with default risk higher than 7%?

5. [5 Points] In 2008, the United States lead the world into a global financial crisis resulting in the 'Great Recession'. The main source of this crisis was that there were too many home mortages that entered into default that it lead to financial institutions going bankrupt. Some argue that the root cause of the problem was that there was too light government regulation of US banks, including Freddie Mac. Others argue that the root cause of the problem was that the government subsidized the mortage industry too much and this lead to excessive risk taking by banks.

A. Explain how machine learning could help banks like Freddie Mac and the US government avoid such crises in the future.

B. Using your analysis above, explain how both the lack of regulation and government subsidies of the mortage market could contribute to such a financial crisis.