DA/CSC 423/523

Final Project

OneBank launched a program to encourage its existing customers to borrow via a consumer loan program. The data are available in file bank.csv. The bank wants to develop a model to predict which customers have the greatest probability of accepting the loan given the data from a previous promotion campaign, to reduce promotion costs and send the offer only to a subset of its customers.

You will develop three models of your choice one which should be Naïve Bayes, then combine them in an ensemble. You will need to preprocess the data, perform feature engineering, partition it into training and validation sets, and fit the three models. Use "loan" as the outcome variable. Report the validation confusion matrix for each of the models. Create a data frame with the actual outcome, predicted outcome, and each of the models. Display this data frame.

Add two columns to this data frame for (a) a majority vote of predicted outcomes and (b) the average of the predicted probabilities. Using the classifications generated by these two methods derive a confusion matrix for each method and report the overall accuracy.

Compare the error rates for the three individual methods and the two ensemble methods outlined above.

By submitting your solution for this assessment, you state "I affirm that I have neither given nor received unauthorized help on this assignment, and that this work is my own."

Collaboration is not allowed; suspected plagiarism will be penalized.

Please note:

- Late submissions will be penalized at 20% for each day it is late.
- Submitting your notebook in incorrect format and then providing the one in correct format after the deadline is considered late submission. Make sure that you submit it in the correct format that I can view and run on my computer after I download it.
- Submissions two days after the deadline will receive no credit.
- Explain each step in your notebook in Markdown.
- You should follow the examples given in the slides/textbook for your code. Obscure code will receive no credit.