Binary Classification by SVM and Logistic Regression

Imagine that you are working on churn prediction project in a bank, and the business owner asked you to send him a list every month with the names of the customers who will probably churn (leave the bank).

Since the business owner receives monthly reports about the number of churners, he already knows that every month the bank loses about 1300 customers.

The business owner illustrated that he would like your list to capture most of the churners (at least 75% of the churners) to decrease his opportunity loss as much as possible.

However, the actions which will be taken based on your list will be very costly, that's why he requires that your list does not exceed 1500 customers (to decrease the expenses on false positives).

Construct a machine learning model that fulfills the previous requirements, following these guidelines (the previous requirements should be fulfilled on the validation data):

1) Use logistic regression and find the best threshold for prediction that fulfills the business requirements using the ROC curve output.
2) Use SVM with all these kernels:
    a. Linear kernel SVM (construct a loop that finds the best C hyper-parameter)
    b. Polynomial Kernel SVM (construct a loop that finds the best polynomial degree d, the best C, and the best Gamma Hyper-parameters)
    c. RBF Kernel SVM (construct a loop that finds the best C, and the best Gamma Hyper-parameters)

    Note that: we decide the best C and the best Gamma based on the best performance on the validation data (the best performance must fulfill the above business requirements)
3) Plot the following curves for the RBF kernel SVM:
    a. Curve demonstrating the effect of changing C on the F1-score of the train-data
    b. Curve demonstrating the effect of changing C on the F1-score of the validation-data
    c. Curve demonstrating the effect of changing Gamma on the F1-score of the train-data
    d. Curve demonstrating the effect of changing Gamma on the F1-score of the validation-data
4) Compare the best setting of Logistic Regression Model with the best setting of SVM Model and use the best model to release a list of names of users who will probably churn (using your validation data)
5) Summarize the best model performance for the business owner (precision, recall, f1-score for each class)