Comments for Spamham.py

You have gone about preprocessing data and handling non numerical values effectively. You have considered logistic regression and support vector Machine only. There are many classification models that could be considered to handle this problem. You may want to consider ensemble techniques like Gradient boosting, Ada boosting and random forest. You are likely to get better results when you use these classification techniques. You may also want to consider voting classifiers that use logistic regression and support vector machines

Please call out if you had specific reasons to choose support vector machines and logistic regression only