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 o**nly**

**Comments dated 23rd may 202**0

Comments for MPS prediction

It looks like you are trying to predict continuous variables using a randomforest regressor. Have you evaluated all other regression options namely linear regression, multi linear regression, polynomial regression, ridge and Lasso . Is there a specific reason why decided to go with random forest regressor?

Looking at the model that you are trying to predict, it looks you are trying to get a relationship between mail per second, test description, model and version did you check for the correlation between these features? How would you know the MPS is a function of test description for instance? Please try out the feature engineering technique that we are talked about like recursive feature elimination