```
% HW4 Q1
% Dong-Bang Tsai

clear;

N_sample = 4601;
Y=zeros(N_sample,1);
Y_type = -1*ones(N_sample,1); % 0 is training set, 1 is testing set.
X=zeros(N_sample,58);

scanformate = '';
for i=1:59
    scanformate = strcat(scanformate, ' %f');
end

fid1  = fopen('spam.data');
fid2 = fopen('spam.traintest');
tline1 = fgetl(fid1);
tline2 = fgetl(fid2);
for i=1:N_sample
    C = textscan(tline1,scanformate);
    for j=1:58
        X(i,j) = C{j};
    end
    Y(i) = C{58};
    C2 = textscan(tline2, '%f');
    Y_type(i) = C2{1};
    tline1 = fgetl(fid1);
    tline2 = fgetl(fid2);
end
fclose(fid1);
fclose(fid2);
clear('fid1','fid2','tline1','tline2','C','C2');

indx_train = find(Y_type == 0);
Y_train = Y(indx_train);
X_train = X(indx_train,:);

indx_test = find(Y_type == 1);
Y_test = Y(indx_test);
X_test = X(indx_test,:);

[Xp_train, Xp_test] = preprocessing(X_train, X_test);


Yest_train = NaiveBayesclassify(Xp_train,Y_train, Xp_train);
train_error = sum(Yest_train ~= Y_train)/length(Yest_train)

Yest_test = NaiveBayesclassify(Xp_train,Y_train, Xp_test);
test_error = sum(Yest_test ~= Y_test)/length(Yest_test)


```