My Final Report

By: Yara Iskandarani
Topic : spam detection

1. The problem statement
   The problem statement is that the data set I had needed to be
   classified . What do I mean by that ? so we have a bunch of
   emails and we don't know whether or not they are ham or spam .
   so i use the rainforest classification algorithm to classify
   all the emails and put then under the category ham of spam

2. explain the data set
 So as i said this data set consists of a bunch of emails and they
need to be put under the category spam of ham so that's what i did

3. why do you need AI to solve this problem
 I need AI to solve this problem because without it we wouldn't
be able to classify whether or not these emails were spam or ham

4. so The idea behind a **random forest** is to combine many
decision trees into a single model. Individually, predictions
made by **decision** trees (or humans) may not be accurate, but
combined together, the predictions will be closer to the mark on
average.

5.
First cell: the first cell of code is just importing things like
numpy and pandas
Second cell:then i imported standard scaler
Third cell:next i printed the columns to see what im working
with
Fourth cell:i added in the iloc values

Fifth and seventh cell:i label encoded x and y to convert to strings

Sixth cell: i reshaped some code. So basically instead of having multiple columns and one list in each . we reshaped them to be one column with all the lists

Eighth cell : next i trained the model

Last cell: then i used all the training to take an email and predict whether or not it is spam or ham

Thank you