**3.THE PROBLEM**

Here,the problem is to obtain a decision function f,that can tell us whether a given email is a spam or non-spam.

**3.1 WHAT ARE SPAM EMAILS?**

Spam emails are the nearly identical emails sent to a bulk of people. These emails sometimes contain some links which may direct the recipient to some phishing website or to the sites that host malware. Spam emails also include attachments that may be harmful to the computer. Spammers collect the email addresses various chatrooms, customer lists, websites .These may also include some offers at various commercial websites which are sent in bulk to all the users who have signed up on that website. Most of the times these spam messages includes images in their mails so that they don’t get blocked by the word filters.In fact ,when a viewer views the image,it is directly downloaded from the spammer’s website so now the spammer will know exactly which users have viewed the image so that they can spam those email addresses more.

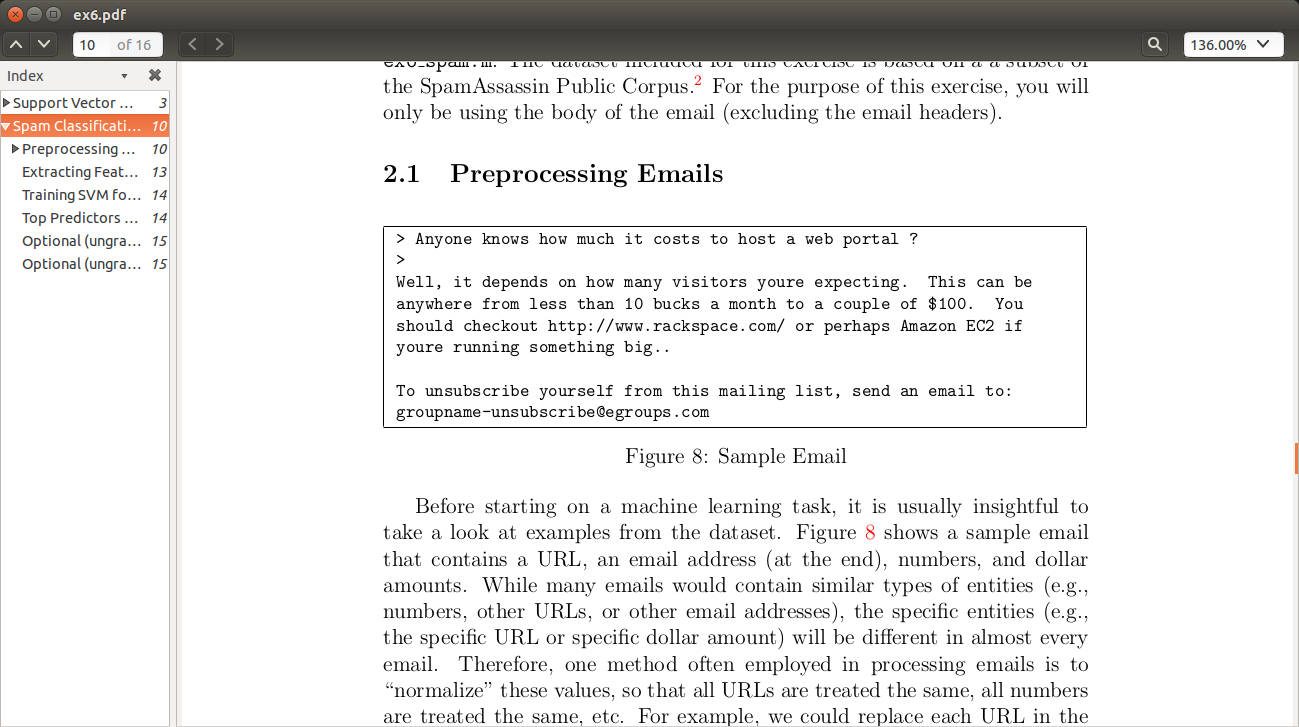


Fig 3.1 A SAMPLE SPAM EMAIL

**3.2 HOW DO WE CLASSIFY EMAILS AS SPAM OR NON SPAM?**

Since these unwanted emails can cause harm to our computers and even more than that,so we need to classify the inbox emails as spam or non-spam so that we don’t open the spam email by mistake.

We train our machine to classify the emails by inputting lot of emails.The machine processes those emails by creating a decision boundary and learn from those emails. Then we test it for some dataset.

Now the machine the same decision boundary to classify the emails.If we get good results for the dataset then we can test the model for our own emails.