

Requirements :

1. Define your feature, e.g. : email_length (does it contains special chars, or plain English)
2. Make good cleaning, make sure nothing else is left
3. Keep duplicate rows, for rows with empty subject/body
 - a. Drop that rows
 - b. Placeholder <Empty>
 - c. None

Division of works :

1. Plain English Yue
2. French xinyi
3. Mars Language Feiyi
4. email + url + IP + HTML + JavaScript zixia
5. Suspicious/dangerous set qianguiqiaoying
6. Money, serial number, tracking number, card number, postal code Guo

Input: plain text

Fn

Output: features

#email, length

#email/length <- not for now

More links

Has javascript related tags: script>

Has HTML tags

Action words: click on a link, fill out a form, submit detailed information

Other words: Paypal, bank, account