

If we have 500 examples in a cross validation set. ($m_{cv} = 500$)

Suppose, the algorithm misclassified 100 of those.

We'll manually examine the common traits in those 100 examples and categorize them.

We might have these common traits:

- Pharma 21
- Deliberate misspellings (w4tches, med1cine) 3
- Unusual email routings 7
- Steal passwords (phishing) 18
- Spam messages embedded in images 5

This error analysis particularly tells us about the major areas where our algorithm messes up so that we can focus on them instead of focusing on areas that don't need that much attention

for eg. pharma and phishing cases are in abundance and we should take care of them instead of wasting time on fixing deliberate misspelling cases.

One email can be classified into many categories and isn't exclusive to only one category. eg. Email 1 can belong to deliberate misspellings and phishing.

- Pharma 21
- Deliberate misspellings (w4tches, medicine) 3
- Unusual email routings 7
- Steal passwords (phishing) 18
- Spam messages embedded in images 5

If our data size is very large (≈ 2000) and 500 of them are misclassified then it's very hard to categorize them by hand.

In that case, we'll randomly take a subset of those 500 cases and categorize them. The size of the subset will be around 100 or a couple hundred.

Q. How to act on those errors?

A. Suppose the maximum errors come from pharma. One approach would be to collect more data, but not just any data. We need more data regarding the pharma spam mails so that our model can train on it and analyze it better.

Another approach would be to **accommodate** **more features** like specific names of drugs, products, etc.

We can also make changes to the algorithm by writing better code. Like for eg. in the case of phishing which is the second most case, we'll write some code which checks the URLs in the email and accommodates some features to check if this URL is suspicious or not?

One limitation of error analysis is that if there are some things which even a human being can't classify then we can't do anything about it.

→ classifying whether an email is spam or not is easy for a human, but stating what ads a person might click on is not possible as it will vary from person to person.