Introduction to Machine Learning

Course Overview

- What is Machine Learning

- When to use ML

- How to use ML

Real-life problem - Spam Detection

- Classify email as Spam or Ham

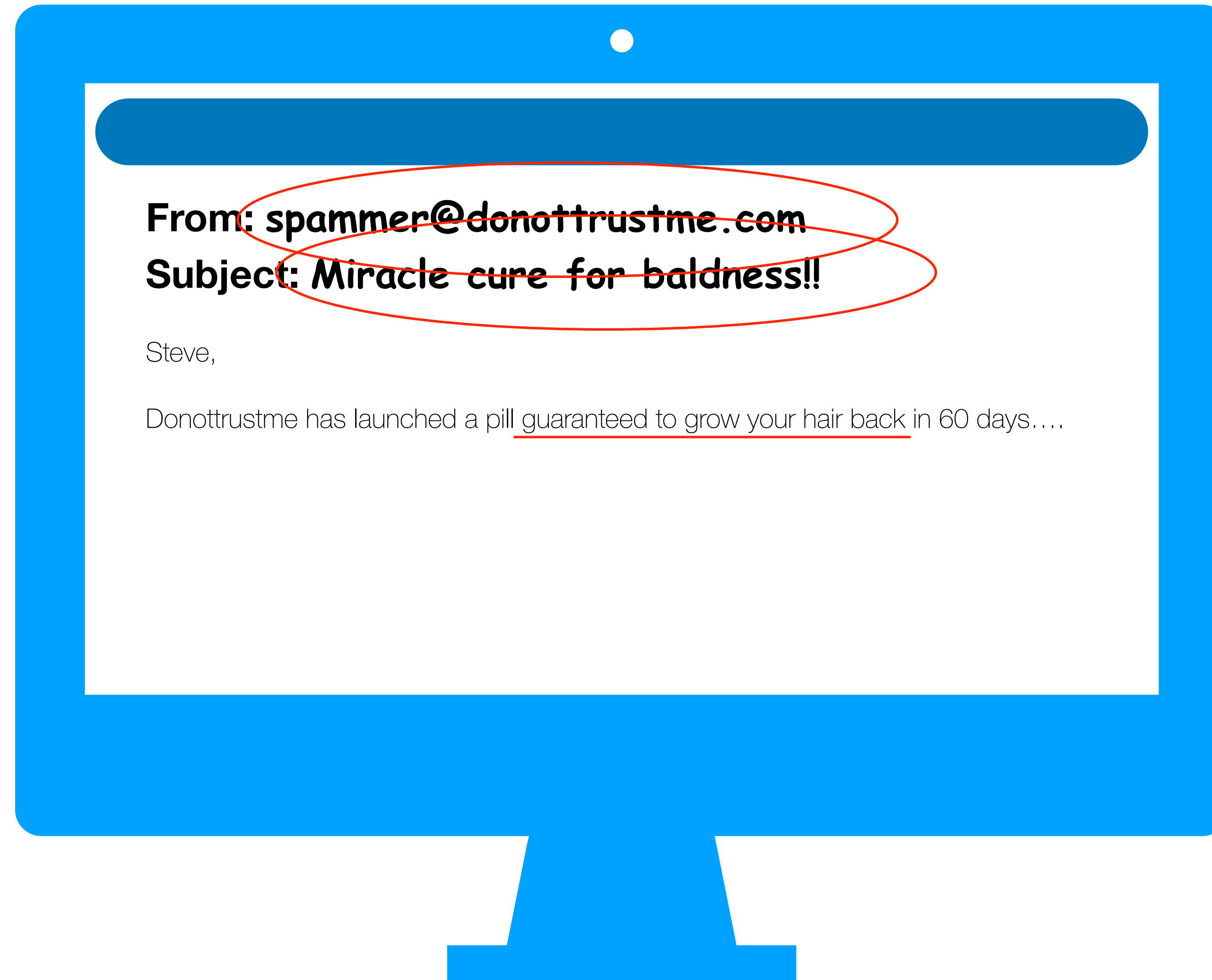- Figure this out before email is sent to inbox

Spam Detection - Rule-based approach

- Define a set of rules to identify spam

- Rules can be intuitive or logical

- Requires domain knowledge
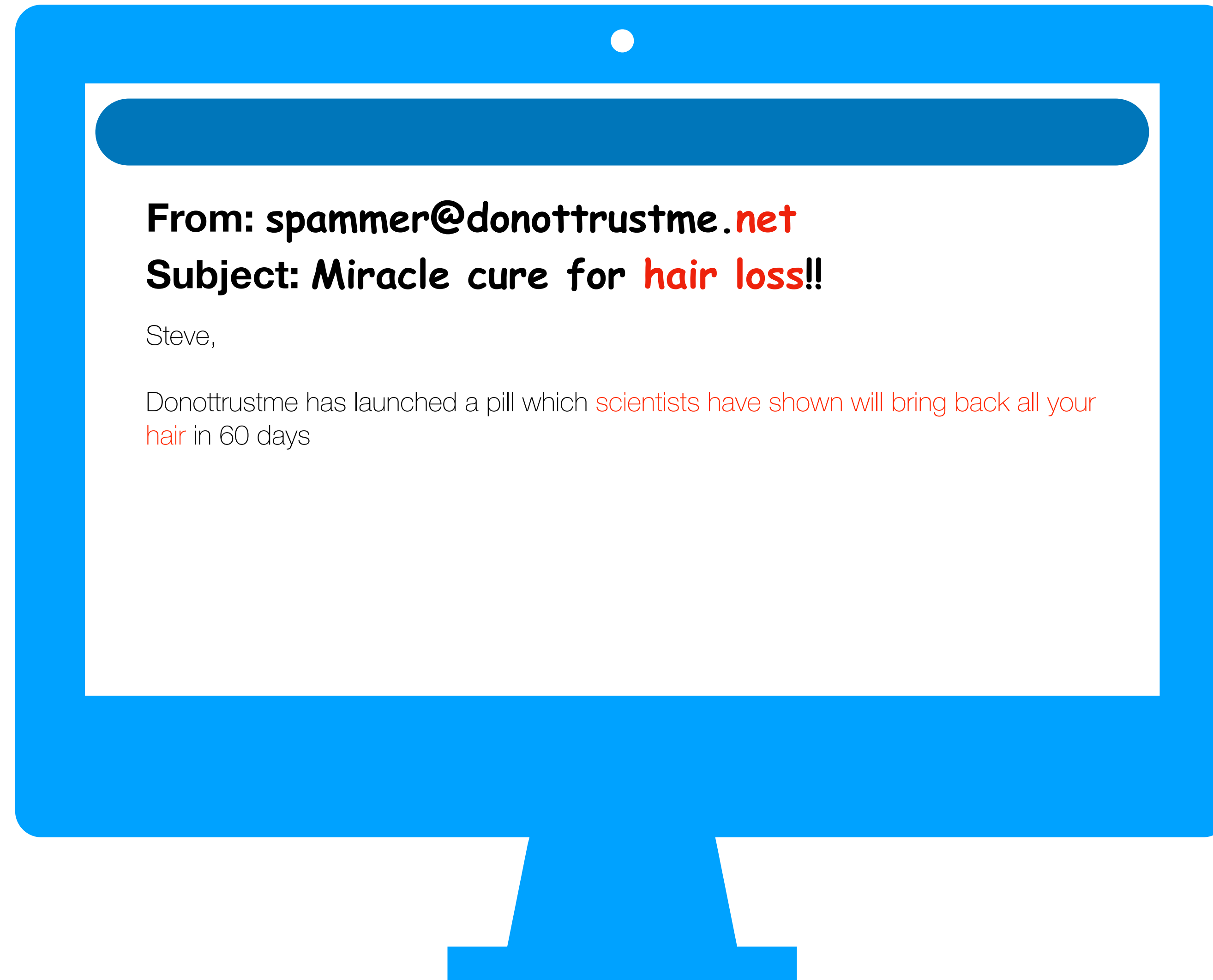
- Need access to historical data

Spam Detection - Rule-based approach

- Blacklist

  - Emails from specific IP addresses
  - Specific words and phrases

- Whitelist

  - Emails from contacts or contact of contacts
  - Certain domains

Spam Detection - Rule-based approach

From: spammer@donottrustme.com
Subject: Miracle cure for baldness!!

Steve,

Donottrustme has launched a pill guaranteed to grow your hair back in 60 days….

Spam Detection - Rule-based approach

From: spammer@donottrustme.net

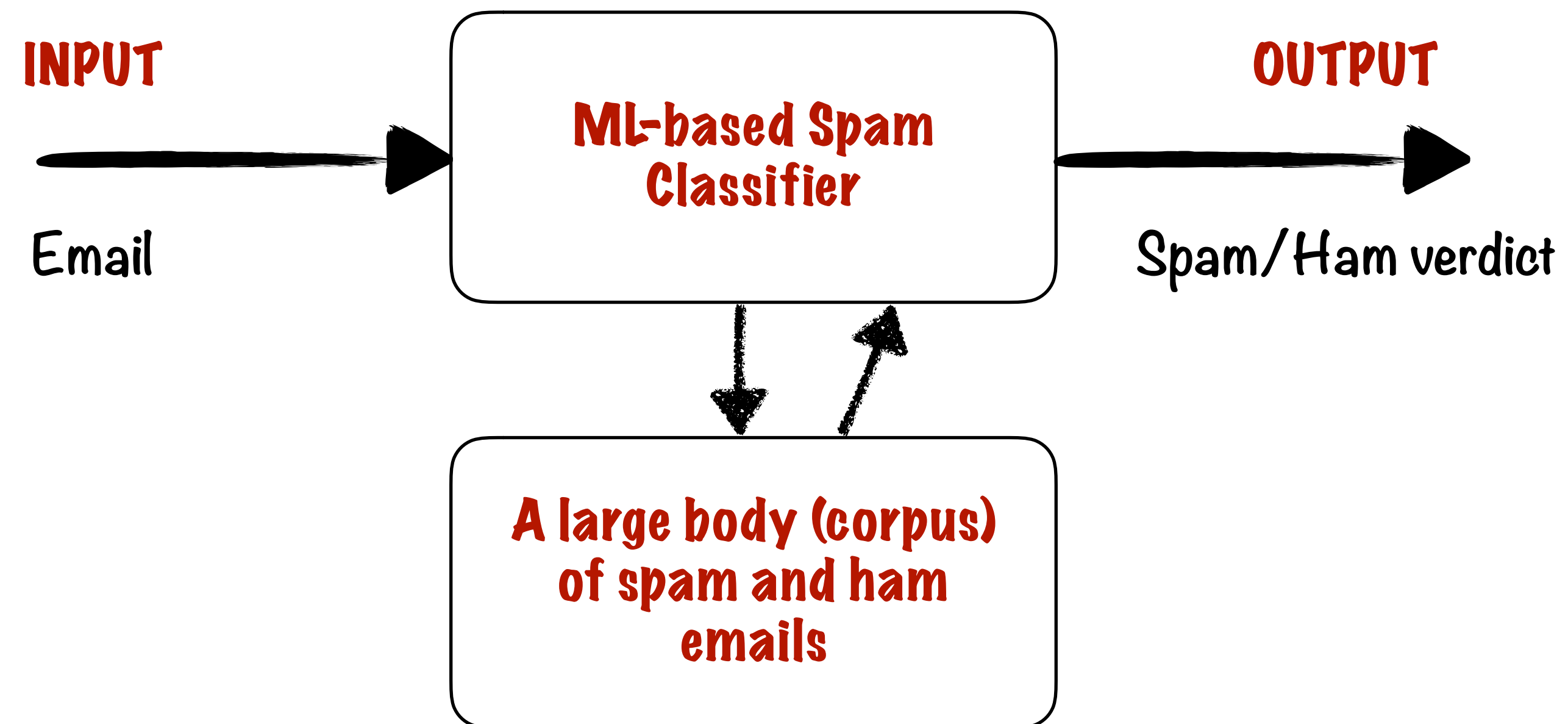**Subject: Miracle cure for hair loss!!**

Steve,

Donottrustme has launched a pill which scientists have shown will bring back all your hair in 60 days

Problems with Rule-based approach

- Rules get outdated

- Rules are static - spammers are dynamic

- Hard to maintain many rules

- Re-thinking rules is time consuming

Spam Detection - ML-based approach

INPUT

Email

ML-based Spam Classifier

OUTPUT

Spam/Ham verdict

A large body (corpus) of spam and ham emails

Spam Detection - ML-based approach

- Algorithm can vary approach based on data

- Corpus can be updated based on user feedback

- Can detect patterns not apparent to humans

- May be less complex than rule-based approach

What is Machine Learning?

Machine Learning is the ability of a program to learn from data

When to use Machine Learning

- The alternative is a complex set of hard-to-maintain rules

- Dynamic environments with changing data

- Rule-based solutions cannot solve the problem

- To get insights on your data

ML-based vs Rule-based

ML-based	Rule-based
Dynamic	Static
Experts optional	Experts required
Need corpus	Corpus optional
Training required	No training