Company ABC

Problem Statement

Company ABC offers a set of self-signup product plans that allow customers to send millions of emails for under a thousand dollars a month. Despite requiring a credit card to activate these plans, bad actors still attempt to abuse the ABC platform to send unsolicited (spam) emails. It is important to catch these spammers as quickly as possible.

ABC’s Compliance Team works diligently to identify and terminate the accounts of these bad actors.

The team suspects that there may be a way of determining if a customer is a spammer by looking at the customer’s email sending practices, how ISPs (such as Gmail, AOL/Yahoo etc) respond, and how engaged recipients are (whether recipients open emails).

Based on 30 days’ worth of data is it build a classifier to identify whether the account is genuine or spammer.

Dataset Description

The dataset is present. The data attribute definition are as follows-

* customer\_id: unique customer id
* rcpt\_domain\_legacy\_isp: the number of emails sent to recipients at legacy ISPs ('aol.com', 'verizon.net')
* rcpt\_domain\_other: number of emails sent to recipients at all other ISPs / organizations etc
* total\_injection\_count: the total number of emails sent to recipients
* total\_injection\_count\_tracked: the total number of emails sent to recipients that had engagement tracking enabled (a subset of total\_injection\_count)
* complaint\_count: the number of spam complaints received
* trap\_count: the number of emails sent to spam\_traps
* open\_count: the number of emails that were opened (a subset of total\_injection\_count\_tracked)
* xxx\_bounce: Different types of bounce.
* terminated: has this account been terminated by ABC’s compliance team?