

Predict email opens

Manuel Felipe Pineda

Universidad Tecnologica de Pereira

Abstract

In this lab I will try to train a system with given metadata for emails sent to users in order to predict whether or not a future email will be opened for each user.

This metadata contains specific information about:

- The user the email was sent to.
- The email that was sent.
- The user's reaction to the email, including (among other things) whether or not the user opened the email.

This metadata was taken from the emails sent to hackerrank users over a certain period of time.

The project is based in a real competition that can be found [here](#)

Keywords: Data Science, Machine learning, Complicated

1. split data

random

2. model validation

- differences
- F1 score

3. strategies

3.1. Random

- score with prob 0.5: 48625 differences: 0.497 % per ok
F1 score 0.392

- prob 0.7: 54999 differences: 0.431 % per ok
F1 score 0.447

3.2. counting

53569 differences: 0.445 % per ok
F1 score 0.443