Data&Data Classification Challenge

Introduction

Our solution tracks counterfeit and grey market activity online. For both cases, classification algorithms are essential to process the large amounts of data generated by our scrapers.

One of the most important steps is to determine if a post on a social media website is selling a product or not, and if it is indeed selling, whether that product is fake or genuine.

The goal of this challenge is to train a classifier capable of classifying Facebook posts. We are providing you a set of labeled data (link below) for training and validation. Feel free to choose any data you want from the training set for your model.

Objective

Your classifier should be able to classify Facebook posts into three categories: fake seller, reseller, no seller. There is no limit on the number of models.

This challenge is open-ended. You are invited to complement your solution with any enhancements that can help highlight your strengths: comparing results, visualization, etc.

Constraints

Our only constraint is that you should use Python, ideally version 3. Beyond that, use any tool that can help you get the job done better or faster.
While this is not mandatory, we recommend you try to stick to the PEP 8 style guideline for best readability.

Compatibility

Your code should be easily deployable. We need to be able to run your code ourselves, preferably in a Unix environment. Leave us instructions for installing any dependencies and getting your script up and running.

Submission

Send us your source code or, preferably, repository URL at jobs@data-and-data.com. An explanation of how you approached the problem, the rationale behind key design choices, and how your solution works (e.g. an architecture overview), is also much welcome.

Submission

Requests for help or advice are welcome. We are more interested in gauging your potential rather than your current technical skills. Not knowing is fine, but learning quickly is essential.

Good luck!

The Data&Data team

Link

https://www.dropbox.com/s/88eexrb3iyctz32/Data%26Data%20Classification%20Challenge%20-%20Facebook%20-%20Training%20Set.zip?dl=0