Ai SANDBOX

ACCELERATE YOUR INNOVATION IN 72HRS

SANDBOX 1

Problem Statement: Body mass index (BMI) prediction from face images

#bmi, Detail: Human faces encode plenty of useful information. Recent studies in psychology and human perception have found that facial features have relations to human weight or body mass index (BMI). These studies focus on finding the correlations between facial features and the BMI

Primary References: http://csee.wvu.edu/~gdguo/myPapers/B...ce2013.pdf

Dataset: http://www.faceaginggroup.com/morph/ (these are paid datasets)

Mentor: Karthik R

Join Slack Channel: https://aisandbox.slack.com/messages/C4NGDCFQV/details/

SANDBOX 2

Problem Statement: Sentiment Analysis for News and Blogs

#sentimentanalysis, Detail: News can be good or bad, but it is seldom neutral. Although full comprehension of natural language text remains well beyond the power of machines, the statistical analysis of relatively simple sentiment cues can provide a surprisingly meaningful sense of how the latest news impacts important entities. Such analysis could reveal profound insights into an election campaign, brand perception and even be extended for fake news detection.

Primary Reference: http://icwsm.org/papers/3--Godbole-Sr...Skiena.pdf

Dataset: Attached a sample. Apart from this dataset, we could extract a lot of data from social media like twitter or scrap some popular new and blog web pages like NY Times, CNN, etc. (attached NYTimesBlogTrain.csv)

Attachments: NYTimesBlogTrain.csv
(http://forgeforward.in/aisandbox/NYTimesBlogTrain.csv)

Mentor: Karthik R

Join Slack Channel: **https://aisandbox.slack.com/messages/C4P4E12AJ/details/**

SANDBOX 3

Problem Statement: Find lookalike audience using user persona

#lookalikeaudience, Detail: How often have you found a Facebook ad that grabs your attention and makes you have a second look while you were through your lazy scrolling on those boring posts of your friends? Perhaps, quite often. A lot of Facebook ads are targeted, which means FB know a great deal about you (more than you know about yourself). Now, that might sound like magic but there is a lot of data being churned to come out with such a precise targeted ads. Imagine if you already have few thousand customers in your business and now your business is looking for ways to growing that to 10 times more. What would you do? Maybe run a TV Commercial (too costly) OR e-mail campaign (too low conversion rate and limited to your customer base). This is where finding lookalike users would help. If you have 100 users and think if you could build a Machine Learning algorithm which could identify another 100 users similar to the existing base using a lot of data associated with users.

Primary Resources:

How FB Targeting Works

a. https://facebook.exceedlms.com/uploads...index.html
b. https://facebook.exceedlms.com/uploads...index.html
c. https://facebook.exceedlms.com/uploads...index.html

Targeted Ad Research

a. http://faculty.haas.berkeley.edu/giye...tgtadv.pdf
b. https://dspace.mit.edu/openaccess-diss...21.1/67901
c. http://hiplab.mc.vanderbilt.edu/~zhan...w/p261.pdfd.
https://www.researchgate.net/file.Post...5625098873.

Datasets: The attached dataset includes two tables: data on the interest categories Facebook shows to users and the ad groups its shows to advertisers. ProPublica used this data to show that Facebook tells its users a lot of things it knows about them, but not all the things it's selling to advertisers.

Interest category data was compiled using a Chrome extension, built by ProPublica reporters. The extension showed users the interest categories Facebook assigned to them, and gave users the opportunity to share all of these categories with ProPublica. The data shared did not include any identifiable user information. Through this extension, ProPublica crowdsourced 52,235 unique interest categories.
The second table contains data scraped from the company's ad buying portal. This table shows what audiences Facebook allows ad buyers to target.

This dataset just tells a lot about categories under which ads are classified. In order to actual user persona, you might have to rely on user feeds from twitter and from words used in the tweets, try to identify the user mapping to these categories (attached Facebook Ad.zip)

Attachments: Facebook Ad.zip (www.forgeforward.in/aisandbox/FacebookAd.zip)

Mentor: Karthik R

Join Slack Channel: https://aisandbox.slack.com/messages/C4NGDLGSH/details/

SANDBOX 4

Problem Statement: Face verification from imageset

#faceverification Detail : Aim is to learn a low dimensional representation of image sets, specifically face sets. Each image set will be transformed into a vector of numbers. The

representation should be such that image sets of same person should be close together, while that of different person should be far apart. The distance is computed using L2 norm.

Dataset
https://www.cs.tau.ac.il/~wolf/ytfaces/

References

[1] Jiwen Lu, Gang Wang, Weihong Deng, Pierre Moulin, and Jie Zhou, “Multi-manifold deep metric learning for image set classification,” in Proceedings of the IEEE Conference on Computer vision and Pattern Recognition, 2015, pp. 1137–1145

[2] Parkhi, O.M., Vedaldi, A. and Zisserman, A., Deep face recognition. Proceedings of the British Machine Vision, 2015, 1(3), p.6.

Mentor: Nirmal, Hari

Join Slack Channel: https://aisandbox.slack.com/messages/C4P4EDPQW/details/

SANDBOX 5

Problem Statement: Influence Maximization in Social Media

#infmax, **Detail:** Now-a-days an effective strategy to market a product is via word-of-mouth. Such a strategy along with the power of social media could prove to be disruptive. As often users resort to social media to air their opinion. Such a behavior has a potential to promote a product if handled aptly. Hence, for a company, it is worth to know a set of individuals who are "influential". In simpler terms, this set of user could in principle influence a large number of other users. This problem is formally known as "Influence Maximization".
For a prototype, scrape twitter to get tweets of 5-10K distinct users. Identify a set of 200 users who if tweets about the mobile phones would be able to spread the information to a large number of users.
Reference: https://www.cs.cornell.edu/home/kleinber/kdd03-inf.pdf
Mentor: Tejas & Ashish

Join Slack Channel: https://aisandbox.slack.com/messages/C4PJWEG58/

For assistance kindly reach out to Yogi, Tech Manager, Ai.Sandbox Team
+91 94433 54315 | yogi@forgeforward.in