FAQs

BASIC GUIDANCE

PROBLEM STATEMENT :

Instagram Follower Analysis (Competition ID: AC03)

This competition will require you to collect data from social media about user attributes like existing followers of the user, user bio, user posts, how long the user is active, location, interests etc to predict if the user will follow you back when you are following the user from a neutral account (neutral account is where you don't reveal personal information about yourself like age, gender etc. You only have a bio, dp and some posts).There are plenty of variables you can derive based on above. Your task will be to develop an algorithm that can classify users into likelihood of following you back in 5 days if you follow them. Your model should be able to assign a 0 or 1 value to any user.

Evaluation Criteria: on the date of evaluation (29th June) a neutral account will be created and it will apply your model to follow 300 random users. The number of users whom your model has assigned a value 1 and who follows back in 5 days will be your score.

What you'll learn :Application of classification and social media networks. How to build a basic logistic regression or Random forest model first and then improve it using different other features. You will also learn how modeling techniques can be used, their merits and demerits and some ideas about ensembling in general.

WHAT DATA TO GET?

Understand the problem statement and try to formulate an initial approach.
Feature engineering:- Do the necessary research about the features that will help you predict if the user will follow you back when you are following the user from a neutral account.(Feature engineering is the process of using domain knowledge of the data to create features that make machine learning algorithms work.)

For example:-

Google Scholar can be searched (with Keywords Instagram; social media; cultural analytics etc) for relevant papers

Some links that might help:-

https://www.forbes.com/sites/tomward/2017/07/21/you-want-more-instagram-followers-heres-how-to-do-it/#4a86732066ae

https://www.quora.com/Why-and-how-do-unknown-people-follow-me-in-Instagram-even-if-I-dont-upload-pictures

https://www.cpcstrategy.com/blog/2018/11/free-instagram-analytics-tools/

Research is considered unnecessary and boring but don't ignore it. Choosing the right features can be the key to better results.

HOW TO GET THE DATA?

Data can be scraped or an instagram API could be used.

1.https://www.youtube.com/(using bs4)

2.https://www.youtube.com/(using bs4)

3.https://www.youtube.com/(using API)

4.https://www.youtube.com/ (A good playlist to learn scraping in general, using bs4 and API)

https://github.com/facebookarchive/python-instagram

https://www.instagram.com/developer/endpoints/

For example(API):-

You should definitely check out Instagram's API, which can provide you all the public information you would want to scrape. You'll just need to write a script to make the proper API calls

https://stackoverflow.com/questions/41405697/how-to-extract-instagram-data?rq=1

Other references:-

https://www.promptcloud.com/blog/how-to-scrape-instagram-data-using-python/

To get number of followers/following (one of the doubts raised):-

https://codepen.io/vivekradhakrishnan/pen/CoFdz?editors=1010

WHAT AFTER GETTING THE DATA?

Data pre-processing

Data pre-processing is one of the most important steps in machine learning. It is the most important step that helps in building machine learning models more accurately. Data pre-processing is a process of cleaning the raw data

Researching the model that will be best for the type of data

Find your model. This is a classification problem (when the target variable is categorical , belongs to either Class A or B) whether the user follows you back or not.

These some most used classification algorithms.

- K-Nearest Neighbor
- Naive Bayes
- Decision Trees/Random Forest
- Support Vector Machine
- Logistic Regression

Testing has to be done at your own level before the evaluation. You can come up with your own ideas for that. The main goal is learning and innovation.

This document is only an initial example reference for those of you who were confused and didn't know how to proceed.

Some of the links have to be pasted to search engine as they don't have the hyperlink. You have to be your own judge for their correctness, relevance and usefulness. You can find your own better references and work accordingly.

All further doubts technical and non-technical are more than welcome. Feel free to contact the SMEs. ALL THE BEST.