**Data File:** instagram-fake-spammer-genuine-accounts.zip

**Problem Description:**

Fake accounts create a big problem for social media platforms. Analyze this data set to suggest how Instagram could detect fake accounts and block them.

Can you create distinct clusters that work for this problem?

Which classification method seems to work best?

**Note:** The zip file contains separate test and training data csv files. I recommend combining them into a large data set and then iterating across a number of trials to randomize the training data.

### File Description:

### Context

Fakes and spammers are a major problem on all social media platforms, including Instagram. This is the subject of my final-year project in which I set out to find ways of detecting them using machine learning. In this dataset fake and spammer are interchangeable terms.

### Content

I have personally identified the spammer/fake accounts included in this dataset after carefully examining each instance and as such the dataset has high level of accuracy though there might be a couple of misidentified accounts in the spammers list as well. The dataset has been collected using a crawler from 15-19, March 2019.

### Inspiration

This dataset could be further improved in quantity and quality measures, but how much accuracy can it achieve?