**Google Playstore Fake Reviews / Opinion Spam detection**

* **Following are the problems and techniques discussed in the papers:**
* **Problem**: It is very hard to identify the gold-standard fake reviews because such reviews are written by humans and in such a way that they are not recognizable by inspection.
* But in context to our problem, we can get gold-standard fake reviews from:
  1. **Makingmoneywithandroid** forums: because we are certain that reviews posted by the forum users are definitely fake reviews and can serve as gold-standard reviews.
  2. **Fiverr:** We can make some honeypot apps and post them on google play, and then ask fiverr seller to post reviews on the honeypot apps.
  3. **Amazon Mechanical Turk:** Post our honey pot app link as HIT and ask the turkers to write reviews in exchange for money.
* **Problem**: How to do feature extraction? Papers show that N-gram models, LWIC, Frequency distribution of POS work well.
* We can extract N-gram models, LWIC, Frequency distribution of POS features from the text of play store reviews, along with:
  1. Metadata of the user’s google plus profile.
     + Number of friends.
     + Number of posts in his feed.
     + Ratio of number of spam posts in his feed.
     + Bio information.
     + Has profile photo.
     + Followers.
     + Page Views, etc.
  2. Metadata of the app for which the review has been posted. These app might tend to have some specific characteristics like:
     + Category of the app.
     + Words in the app description.
     + Count of links in app description.
     + Rating of the app.
     + Number of apps posted by the developer.
     + Is developer information complete?
  3. Time distribution of the reviews posted by the same user, like, he reviews in a burst or evenly.
* **Problem**: One of the paper claimed that if a user posts duplicate reviews for various products then that user is a fake reviewer.
* We could have applied the same technique here, but as of now it is not possible to get all the reviews posted by a single user. We can only get the rating given by him in the past, not the text of the review. If we some get this information, we can do duplicate checking.