

Annotating Forum Posts

Thank you for helping me create an annotated dataset for my dissertation project.

You will have a csv file containing the forum posts, there are two blank columns called Label and Sentiment. For each post you need to assign the best fitting category and the sentiment of the post.

Labels

For each post please enter just the number of the best fitting category into the 'label' column.

For quick reference these are the labels with their associated number. Please see below for more detailed explanations:

0. Mistakenly filtered
1. Sex
2. Relationships
3. eWhoring
4. Online crime and harassment
5. Descriptions
6. Real world abuse
7. Politics and ideology
8. Skills in technology
9. Story

0. Mistakenly filtered

- A spelling mistake has caused the post to be selected it does not actually refer to women in any context
- Links that have been removed may have contained words that were picked up in the filtering

1. Sex

- Any post talking about sexual experiences (fictional or not), porn, contraceptives
- Including jokes or insults about sexual experience

2. Relationships

- Asking for or giving advice on dating and relationships
- Talking about friends and family
- Posts about texting or talking to women
- Talking about breakups or experiences of relationships

3. eWhoring

- eWhoring is a fraud technique where (usually) men pretend to be women online where they flirt and chat to men, earning their trust before selling them sexually explicit photos and videos. Often the photos have been stolen and are distributed without consent from the owner.
- Posts will mention selling or buying "packs" which are collections of images
- Social engineering (SE) advice on how to impersonate women

- Talking about different dating sites with lots of men on
 - Talking about verification images, such as how to photoshop into specific poses or add pieces of paper into pictures, as well as asking female members of the community to send/sell voice recordings or talk on the phone
 - Advice on fake bank, PayPal or other money related accounts
 - Posts about stealing nudes or obtaining nudes through hacking or the use of RATs (remote access trojans)
 - Buying and selling fake facebook or instagram accounts
 - Buying and selling (girl) slaves, referring to access of webcams
4. Online Crime and Harassment
- Mentions of hacking to obtain information or for fun
 - General social engineering techniques to gain passwords or private information
 - Including offering to sell hacking services
 - Posts asking for advice about hacking
 - Mentions of use of bots
 - Any form of abuse and harassment performed online, for example “revenge porn”, denial of service (DoS) attacks, blackmail as a result of hacked information or images
5. Descriptions
- Descriptions or qualities of women, real and fictional, including appearances and personalities
 - Including discussion of preferences in women
6. Real world abuse
- Derogatory comments
 - Discussion of physical abuse and violence
 - Harassment that takes place offline, eg physical stalking and following
 - Mentions of sexual abuse including underage girls
7. Politics and Ideology
- Political discussion and opinions
 - Human rights and equality
 - Discussions on religion and philosophy
 - Current affairs and events in the media
8. Skills in hacking, gaming and technology
- Discussing women in the context of their ability or interest, including general remarks and specific examples
9. Story
- The writer is telling a story that happens to mention a woman
 - Not an opinion and not within any of the other categories
 - The writer is addressing the community, eg “hello ladies and gentlemen”

Sentiment

Each post also needs to be labeled with the sentiment towards women being expressed.

Fill in the “Sentiment” column with the following values:

- P for positive

- N for negative
- 0 (the number zero) for neutral (this default value is already filled in)

Examples of negative sentiment will include posts about hacking womens accounts or devices, mentions of abuse and derogatory statements. Positive sentiment includes posts dissuading the use of hacking against women, advocating equal rights and respect towards women and other positive opinions about women.

We are looking at the opinion expressed from the view of society (ie crime is bad) rather than from the writer. There may be cases of the writer expressing an opinion such as "i love hacking into girls snap accounts", from their perspective this would be positive, but in this classification task the act of hacking is negative so this would be labelled as negative sentiment.