**Dataset Field (A to Z ascending)**

|  |  |
| --- | --- |
| **“alias”** | alias name shown under the profile picture |
| **“date”** | timestamp of each post |
| **“description”** | caption of each post |
| **“descriptionProfile”** | user biography/introduction |
| **“filename”** | Local image accompanied each post.  Will be used for loading images in “cnn/load.py”.  Multiple images of each post are separated by comma. |
| **“isVideo”** | Post type. Null for no image; true for video post; false for image post. |
| **“localization”** | Location for where the post occurred. (Better change to location) |
| **“mentions”** | account mentioned in caption |
| **“multipleImage”** | if post contains multiple images |
| **“numberFollowers”** | number of followers (each user) |
| **“numberFollowing”** | number of accounts this user’s following |
| **“numberLikes”** | number of likes of each post |
| **“numberPosts”** | total number of posts each user has till the time of crawling this dataset |
| **“tags”** | hashtags in the caption |
| **“url”** | url of each post |
| **“urlImage”** | url of image |
| **“urlImgProfile”** | url of profile image |
| **“urlProfile”** | url of each user |
| **“username”** | username(shown on the top) |
| **“website”** | website specified by each user |

**Data Preparation**

|  |  |
| --- | --- |
| **“alias”** | none |
| **“date”** | remove year 1970 |
| **“description”** | TODO: Remove foreign language words? |
| **“descriptionProfile”** | none |
| **“filename”** | none |
| **“isVideo”** | remove null/true |
| **“localization”** | none |
| **“mentions”** | none |
| **“multipleImage”** | none |
| **“numberFollowers”** | none |
| **“numberFollowing”** | none |
| **“numberLikes”** | none |
| **“numberPosts”** | none |
| **“tags”** | none (note: need to generate number of tags) |
| **“url”** | none |
| **“urlImage”** | none |
| **“urlImgProfile”** | none |
| **“urlProfile”** | none |
| **“username”** | none |
| **“website”** | none |

