**Data Dictionary of CS+ project**

Database: ‘Helpme’

Collection 1: ‘Training’. This collection stores training label data.

Collection 2: ‘Test’. This collection stores test label data.

Collection 3: ‘Tweets’. This collection stores tweets data.

P.S.

For feature ***media and metadata***, Use this link to check details.

Technically we can get all the features.

<https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object>

Collection 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Training | | | | |
| Feature name | Level | Feature type | Description | Samples |
| **\_id** | 1 | objectid | MongoDB unique ID for each document | 5bcbef1c7cd13e38708347f7,  5bcbef1c7cd13e38708347fa |
| **postID** | 1 | string | Tweets ID. Used to retrieve Tweets | 324702774006775809,  243445671158489089 |
| **categories** | 1 | Array | Used to store categories. (may contain more than one categories) | [0,] |
|  | 2 | String | Categories name. | ContinuingNews  MultimediaShare  Factoid  Irrelevant  Sentiment |
| **indicatorTerms** | 1 | Array | Used to store indicatorterms. (may contain more than one indicatorterms) | [0,]  [0,1] |
|  | 2 | string |  | Photo  Linked  Video  earthquake  Based\_On\_Author  dead |
| **priority** | 1 | string | priority of this tweets | Low  Medium  High |

Collection 2:Test(processing)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | | | | |
| Feature name | Level | Feature type | Description | Samples |
| **postID** | 1 | string |  |  | |
| **timestamp** | 1 |  | date information |  |
| **categories** | 1 |  |  |  |
| **indicatorTerms** | 1 |  |  |  |
| **priority** | 1 |  |  |  |
| **text** | 1 |  |  |  |

Collection 3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tweets | | | | | |
| Feature name | | Level | Feature type | Description | Samples |
| **\_id** |  | 1 | objectid | MongoDB unique ID for each document | 5bcbef1c7cd13e38708347f7,  5bcbef1c7cd13e38708347fa |
| **identifier** |  | 1 | string | Tweets ID. Used to retrieve Tweets | 351734622884855808,  351741648327294977 |
| **text** |  | 1 | string | Tweet content | "#yycflood be sure when you are ready for re building you call companie..." |
| **type** |  | 1 | string | show this is a tweet(useless) | "TWEETS“ |
| **timestamp** |  | 1 | number | Tweet’s unique time format | 1447495137087,  1396431371270 |
| **lang** |  | 1 | String | machine-detected language. | ‘en’ |
| **score** |  | 1 | number | useless, no meaning for us. | 0 |
| **location** |  | 1 | object | location informations |  |
|  | **description** | 2 | string/null | null feature | null |
|  | **geolocation** | 2 | document | store latitude and longitude |  |
|  | **latitude** | 3 | number |  | 45.42178 |
|  | **longitude** | 3 | number |  | -75.69118399999999 |
|  | **radius** | 2 |  | useless, no meaning for us. | 0 |
|  | **identifier** | 2 | null/string | unique string for location | 5ef832bb704339b0 |
|  | **name** | 2 | string | city or nation name | Sydney (New South Wales, Australia)  United States |
| **statisicalData** |  | 1 |  |  |  |
|  | **User Followers** | 2 | number |  | 4867 |
|  | **User Statuses** | 2 | number | number of Tweets of this user | 37281 |
|  | **User Follows** | 2 | number |  | 2190 |
| **media and metadata** |  |  |  | Other features not mentioned. Huge volume. |  |

P.S.

For media and metadata, Use this link to check details.

Technically we can get all the features.

<https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/user-object>