**Panduan Pelabelan Data**

* Ujaran kebencian (*hate speech*) merupakan tindakan dengan tujuan menyinggung individu maupun kelompok tertentu berdasarkan suku, agama, ras dan antar golongan (SARA). Tindakan tersebut memiliki target, kategori, dan tingkat tertentu (Komnas HAM, 2015).
* Ujaran kebencian ditujukan kepada seseorang (individu) atau suatu kelompok (kelompok agama, kelompok ras, kelompok politik, dan lain-lain).
* Kategori dari ujaran kebencian yaitu sebagai berikut.
  + Agama/kepercayaan

Ujaran kebencian atas dasar agama (Islam, Kristen, Katolik, dan lain-lain), organisasi agama, atau suatu kepercayaan.

* + Ras/etnis

Ujaran kebencian berdasarkan ras (golongan bangsa berdasarkan ciri-ciri fisik) atau etnis.

* + Fisik/disabilitas

Ujaran kebencian berdasarkan perbedaan fisik (bentuk wajah, mata, dan bagian tubuh lainnya) atau disabilitas (autisme, idiot, buta, tuli, dan lain-lain).

* + Gender/orientasi seksual

Ujaran kebencian berdasarkan jenis kelamin, mengutuk seseorang (atau suatu kelompok) menggunakan kata-kata yang merendahkan gender (misal: pelacur) atau orientasi seksual yang menyimpang (misal: homoseksual, lesbian, dan lain-lain).

* + Ujaran kebencian lainnya

Ujaran kebencian dalam bentuk ejekan, umpatan atau fitnah. Ujaran kebencian ini tidak berhubungan dengan empat kategori sebelumnya.

* Tingkatan ujaran kebencian yaitu sebagai berikut.
  + Ujaran kebencian lemah

Ujaran kebencian yang ditujukan untuk individu tanpa menghasut/memprovokasi (biasanya masalah pribadi).

* + Ujaran kebencian sedang

Ujaran kebencian dalam bentuk sumpah/penistaan/stereotip/pelabelan yang ditujukan pada suatu kelompok tanpa menghasut/memprovokasi. Konflik yang terjadi diperkirakan terjadi hanya di media sosial.

* + Ujaran kebencian kuat

Ujaran kebencian dalam bentuk sumpah/penistaan/stereotip/pelabelan yang ditujukan pada seseorang (individu) atau suatu kelompok dengan menghasut/memprovokasi sehingga diprioritaskan untuk segera diselesaikan. Ujaran kebencian ini dapat mengundang konflik yang tersebar luas dan menyebabkan konflik/kehancuran di dunia nyata.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Identifikasi Ujaran Kebencian**





Sumber: Buku Saku Penanganan Ujaran Kebencian (Hate Speech) Komnas HAM

**Contoh Twit Berdasarkan Label**

|  |  |  |
| --- | --- | --- |
| No. | Label | Twit |
| 1. | *Abusive* | USER x || Lo ngomong aja kaga bener bangke' |
| 2. | *Hate Speech* | Lengserkan jokowi krn dia bukan saja telah banyak melanggar ketentuan perundang-undangan,tapi juga tak becus kelola negara |
| 3. | *Individual Hate Speech* | Presiden paling buruk dalam sejarah bangsa2 cebong' |
| 4. | *Group Hate Speech* | Pantesan tim sukses Anies-Sandi ini kalo ngetwit ngaco sok jagoan |
| 5. | *Religion Hate Speech* | USER USER Surban putih jubah putih ht nya busuk2 se busuk tae onta bro.' |
| 6. | *Race Hate Speech* | USER USER Etnis cina memang ada yg ikut berjuang tapi lebih sangat banyak yg jadi penghianat jd antek belanda.. |
| 7. | *Physical Hate Speech* | perempuan kaya lo mending mati aja deh, jelek aja, gausa sok jadi make up artist! |
| 8. | *Gender Hate Speech* | Kamu transgender atau gmn anjing :( URL |
| 9. | *Other Hate Speech* | RT USER: Admin tolol di partai tolol\nPartai Tolol untuk orang Tolol' |
| 10. | *Weak Hate Speech* | USER siapa bego lu bukan tmn gua najis dah kesel sendiri gue' |
| 11. | *Moderate Hate Speech* | USER cebong emang jagonya bikin dan sebar hoax' |
| 12. | *Strong Hate Speech* | 4 partai yg walkout adlh partai besar semua. Banteng nyungsep. Kunci ditangan Rakyat.Jgn nunggu smp2019 langsung lengserkan Jokowi. REVOLUSI |

**Instruksi Pelabelan Data**

1. Data yang akan diberi label merupakan dataset yang berisi twit dari berbagai pengguna Twitter.
2. Beri label yang sesuai pada setiap baris data (twit).
   1. Perhatikan kesesuaian konten twit dengan *hashtag* twit. Apabila *hashtag* twit tidak sesuai dengan konten, maka *hashtag* twit dapat diabaikan.

Contoh:

|  |  |  |
| --- | --- | --- |
| No. | Twit | Keterangan |
| 1. | Tipikal bani 212 bin onta kampret : cepat memuji org yg dihormati, cepat juga mencaci nya kembali.... salam TGB ... ngoahaha #BubarkanPKS' | *Hashtag* twit sesuai dengan konten twit. Twit tersebut termasuk ujaran kebenciandengan tingkat “Strong” karena *hashtag* “#BubarkanPKS”. |
| 2. | #PersibDay Kamvio Soccerball best Indonesian Soccerball Product #bola #ShameOnYouMalaysia #bolafutsal #bolatime #BolaSepak #bolanasional | *Hashtag* twit tidak sesuai dengan konten twit. |

* 1. "USER" adalah akun yang di-*mention* dan diasumsikan sebagai satu individu.
  2. Beri label “1” pada kolom “Abusive" jika twit tersebut mengandung kata kasar. Jika tidak, beri label “0” pada kolom “Abusive”.
  3. Beri label “1” pada kolom “HS” jika twit tersebut termasuk ujaran kebencian (*hate speech*). Jika tidak, beri label “0” pada kolom “HS”.
  4. Apabila twit telah diberi label “1” pada kolom “HS”, maka lakukan identifikasi target, kategori, dan tingkat ujaran kebencian pada twit tersebut.
     1. Target ujaran kebencian (*Individual/Group*)
* Beri label “1” pada kolom “HS\_Individual” jika twit tersebut termasuk ujaran kebencian yang ditujukan untuk seseorang (individu).
* Beri label “1” pada kolom “HS\_Group” jika twit tersebut termasuk ujaran kebencian yang ditujukan untuk suatu kelompok. Apabila kolom “HS\_Group” diberi label “1”, maka kolom “HS\_Weak” harus diberi label “0”.
* Apabila salah satu kolom target ujaran kebencian telah diberi label “1”, maka kolom target ujaran kebencian yang lain harus diberi label “0”. Contoh: Apabila kolom “HS\_Group” telah diberi label “1”, maka kolom “HS\_Individual” harus diberi label “0”.
  + 1. Kategori ujaran kebencian (*Religion/Race/Physical/Gender/Other*)
* Beri label “1” pada kolom “HS\_Religion” jika twit tersebut merupakan ujaran kebencian atas dasar agama. Jika tidak, beri label “0” pada kolom “HS\_Religion”.
* Beri label “1” pada kolom “HS\_Race” jika twit tersebut merupakan ujaran kebencian atas dasar ras atau etnis. Jika tidak, beri label “0” pada kolom “HS\_Race”.
* Beri label “1” pada kolom “HS\_Physical” jika twit tersebut merupakan ujaran kebencian atas dasar perbedaan fisik atau disabilitas. Jika tidak, beri label “0” pada kolom “HS\_Physical”.
* Beri label “1” pada kolom “HS\_Gender” jika twit tersebut merupakan ujaran kebencian atas dasar gender. Jika tidak, beri label “0” pada kolom “HS\_Gender”.
* Beri label “1” pada kolom “HS\_Other” jika twit tersebut merupakan ujaran kebencian yang tidak terkait dengan empat kategori di atas. Sebaliknya, beri label “0” pada kolom “HS\_Other” jika twit tersebut terkait dengan salah satu kategori di atas.
* Apabila kolom “HS\_Other” telah diberi label “1”, maka kolom kategori ujaran kebencian yang lain harus diberi label “0”.
  + 1. Tingkat ujaran kebencian (*Weak/Moderate/Strong*)
* Beri label “1” pada kolom “HS\_Weak” jika twit tersebut merupakan ujaran kebencian dengan tingkat lemah. Apabila kolom “HS\_Weak” diberi label “1”, maka kolom “HS\_Individual” harus diberi label “1” dan kolom “HS\_Group” harus diberi label “0”.
* Beri label “1” pada kolom “HS\_Moderate” jika twit tersebut merupakan ujaran kebencian dengan tingkat sedang.
* Beri label “1” pada kolom “HS\_Strong” jika twit tersebut merupakan ujaran kebencian dengan tingkat kuat.
* Apabila salah satu kolom tingkat ujaran kebencian telah diberi label “1”, maka kolom tingkat ujaran kebencian yang lain harus diberi label “0”. Contoh: Apabila kolom “HS\_Moderate” telah diberi label “1”, maka kolom “HS\_Weak” dan “HS\_Strong” harus diberi label “0”.
  1. Apabila twit telah diberi label “0” pada kolom “HS”, maka beri label “0” pada semua kolom berikut.
* HS\_Individual
* HS\_Group
* HS\_Religion
* HS\_Race
* HS\_Physical
* HS\_Gender
* HS\_Other
* HS\_Weak
* HS\_Moderate
* HS\_Strong

1. Silakan lihat berkas “1\_phase\_gold\_standard.pdf” dan “2\_phase\_gold\_standard.pdf” sebagai referensi pelabelan data.