



UNIVERSITAS
INDONESIA

Veritas, Probatum, Institut

FAKULTAS

ILMU
KOMPUTER

PANDUAN ANOTASI

KODE KEBENCIAN (HATE CODE) DARI DATA TEKS TWITTER

DISUSUN OLEH:

DR. INDRA BUDI, S.KOM., M.KOM.

DAMAYANTI ELISABETH

M. OKKY IBROHIM

NUR INDAH PRATIWI

Magister Ilmu Komputer

Fakultas Ilmu Komputer

Universitas Indonesia

2019

A. PROLOG

Saat ini pengguna media sosial semakin kreatif dalam menyampaikan ujaran kebencian. Salah satu tindakan untuk menyamarkan ujaran kebencian adalah menggunakan kode tertentu untuk mengecoh aturan kebijakan yang diterapkan oleh media sosial. Kode yang digunakan dalam ujaran kebencian tersebut adalah **kode kebencian (*hate code*)**. BBC Indonesia telah memberikan laporan berupa berita daring (*online news*) tentang penggunaan kode kebencian yang dibahas oleh masyarakat Indonesia.¹² Salah satu media sosial yang sering disoroti dalam penyebaran ujaran kebencian di Indonesia adalah Twitter.

Ekstraksi informasi dari Twitter yang megandung ujaran dan kode kebencian diperlukan untuk membantu penanganan penyebaran ujaran kebencian di media sosial.³

Pentingnya penanganan ujaran kebencian memerlukan bantuan dan kerjasama dari berbagai pihak, termasuk Anda sebagai *anotator*.

B. TUJUAN ANOTASI

Tugas Anda adalah menentukan kode kebencian untuk tiap twit pada data.

C. DEFINISI

Ujaran kebencian (*hate speech*) merupakan segala bentuk komunikasi meremehkan atau menghina yang berpotensi menimbulkan konflik sosial, tindak diskriminasi, kekerasan, dan/atau penghilangan nyawa suatu individu/kelompok manusia sehingga perlu ditangani secara khusus.

Kode kebencian (*hate code*) merupakan kode yang digunakan untuk menyamarkan nama individu (orang) atau kelompok tertentu dalam menghina, mengejek, memfitnah atau jenis ujaran kebencian lainnya terhadap individu atau kelompok tersebut. Kode kebencian yang digunakan oleh pengguna internet dapat berupa suatu istilah atau julukan yang disematkan pada individu atau kelompok tertentu karena peristiwa tertentu atau yang menjadi ciri khasnya.

Contoh Kode Kebencian dari BBC Indonesia:

No.	Kode kebencian (<i>Hate Code</i>)	Keterangan
1	cebong atau cebongers atau bong	merujuk pada kata ganti kelompok yang mendukung Jokowi
2	kecebong amis atau kecebong	turunan cebong – pendukung Jokowi
3	raja kodok atau Jokodok	merujuk kepada Jokowi

¹ <https://www.bbc.com/indonesia/trensosial-39562532>

² <https://www.bbc.com/indonesia/trensosial-42814349>

³ <https://geotimes.co.id/opini/kecebong-dan-kampret-label-label-buruk-dalam-pemilu/>

No.	Kode kebencian (<i>Hate Code</i>)	Keterangan
4	kampret	merujuk pada kata ganti untuk kelompok pendukung oposisi Jokowi.
5	IQ 200 sekolam	merujuk pada kata ganti kelompok yang mendukung Jokowi dari pihak oposisi
6	kaum kafirun dan munafikun	diambil dari istilah agama. Berasal dari kata kafir (ka-fir / nomina) dan munafik (mu-na-fik / adjektiva), tetapi ditambahkan akhiran -un menjadi munafikun dan dikelompokkan sebagai 'kaum'
7	kaum bumi datar	dalam konteks pilkada atau pilpres, frasa ini digunakan netizen untuk merujuk pada kalangan fanatik agama yang dianggap mudah marah, sulit menerima kritik, dan mudah dimanfaatkan oleh kepentingan politik. Tetapi bila dikritik atau diberi masukan, kelompok ini akan lebih marah-marah lagi.
8	kelas menengah ngehe	biasanya merujuk pada warga berpenghasilan sedang yang banyak maunya dan banyak tuntutan
9	bani taplak atau bani serbet	kata bani di KBII memiliki arti anak; anak cucu; keturunan: -- Adam, seluruh umat manusia. Bani taplak dan bani serbet merujuk para pendukung Ahok (juga dikenal sebagai sebutan Ahokers) yang berarti 'kaum serbet atau taplak.' Kata serbet dan taplak disematkan karena pada kampanye pilkada DKI, ahok dan pendukungnya menggunakan seragam kotak-kotak seperti serbet dan taplak.
10	bani daster atau kaum cingkrang atau bani jenggot atau tato jidat atau bani jidat gosong	merujuk pada kelompok-kelompok Islam yang berpartisipasi dalam aksi-aksi demonstrasi 'bela Islam' di Monas dan mengenakan pakaian gamis ala Arab
11	bani micin atau kaum kebanyakan micin	merujuk pada kelompok yang anti-Jokowi atau kerap menyalahkan Jokowi atas berbagai kebijakan yang diambilnya atau mungkin hal yang, oleh pendukung Jokowi, dianggap tidak ada hubungannya dengan kebijakan Jokowi
12	kaum penthol korek atau kaum sumbu pendek	merujuk pada kalangan fanatik agama yang dianggap mudah tersulut kemarahannya oleh hal-hal yang dianggap menghina Islam atau ulama

Contoh Twit yang mengandung kode kebencian:

No.	Twit dan Kode Kebencian	Kode Kebencian	Alasan untuk Kode Kebencian
1	"USER USER USER pikiran anak bani kampret g beda jauh sma penista agama . Penfitnah, Penipu Dan Garong.."	bani kampret, penista agama	bani kampret : julukan untuk kelompok yang mendukung oposisi dari Jokowi penista agama : julukan untuk tokoh Ahok terkait tindakan penistaan agama
2	"Ahok and Bani serbet and wereng Coklat MEMANCING UMAT MUSLIM supaya marah? dengan #IklanAhokJahat. KALIAN ITU KUMPULAN ORANG SINTING"	bani serbet, wereng coklat	bani serbet : pendukung Ahok karena saat kampanye pilkada DKI mereka menggunakan pakaian kotak-kotak dengan corak seperti serbet wereng coklat : julukan untuk pemerintahan karena seragam pemerintahan yang umum berwarna coklat.
3	"RT USER: USER Ada CEBONG GENDHENG NGAMUK2 belain si HOMO PENISTA AGAMA!! HAHHA USER USER USER "	cebong, penista agama	cebong : merujuk pada kata ganti kelompok yang mendukung Jokowi penista agama : julukan untuk tokoh Ahok terkait tindakan penistaan agama
4	Kenapa hanya Ahok, Sukmawati dan Ganjar Pranowo (walau kecele) yang mereka sebut penista agama karena menguntungkan secara politik. Ahok dalam pilgub Jakarta yang penuh aroma rasis.	-	Meskipun ada frasa penista agama , frasa tersebut bukan ditujukan sebagai ujaran kebencian terhadap Ahok.
5	#artikata: - cebong = anak katak yang masih berinsang - kampret = kelelawar kecil pemakan serangga #etimologi julukan: - cebong: Jokowi didukung anak katak (cebong) karena hobi memelihara katak. - kampret: Prabowo didukung partai dalam Koalisi Merah Putih/KMP (KaMPret).	-	Meskipun ada kata cebong dan kampret , kata-kata tersebut bukan ditujukan sebagai ujaran kebencian terhadap Jokowi dan Prabowo. Akan tetapi, twit tersebut bertujuan untuk memberitahukan makna kata cebong dan kampret.
6	" USER USER USER Mantap!!!!kemajuan yg luar binasa,,, Salam cebong mania"	cebong	Meskipun twit ini bukan twit ujaran kebencian, tetapi memiliki kode kebencian cebong yang merujuk pendukung Jokowi. Dalam hal ini, penulis twit bangga menjadi cebong .

D. PROSEDUR ANOTASI


Anda akan diberikan data excel yang berisi sekumpulan twit. Twit-twit tersebut bisa termasuk ujaran kebencian atau bukan. Anda ditugaskan untuk menganotasi setiap twit apakah twit tersebut mengandung kode kebencian atau tidak, dengan petunjuk sebagai berikut:

1. Jika menurut Anda twit mengandung kode kebencian, maka tulis semua kode kebencian yang ada di twit pada kolom **Kode Kebencian** menggunakan huruf kecil (non kapital). Jika kode kebencian lebih dari satu kode, silakan pisah dengan tanda koma. Jika sama sekali tidak ada kode kebencian, tuliskan '-' (strip).
2. Kode kebencian dapat berupa kata (misal: *cebong*, *kampret*, dan lain-lain) atau frasa (misal: *bani cingkrang*, *raja kodok*, dan lain-lain).
3. Kode kebencian bisa juga ditemui pada *hashtag*.
4. *Username* yang berawalan dengan karakter @ dan memuat kode kebencian diabaikan. Contoh: @jokodok dan @BaniTaplak. Karakter tersebut bukan kode kebencian.

5. Mohon menuliskan kode kebencian tanpa typo.

Contoh data yang telah dianotasi:

No.	isi_tweet	Kode Kebencian
1	"[REDACTED] [REDACTED] pikiran anak bani kampret g beda jauh sma penista agama. Penfitnah, Penipu Dan Garong.."	bani kampret, penista agama
2	"[REDACTED] Mantap!!!!kemajuan yg luar binasa,,,"	cebong
3	"Ahok and Bani serbet and wereng Coklat MEMANCING UMAT MUSLIM supaya marah? dengan #klanAhokJahat. KALIAN ITU KUMPULAN ORANG SINTING"	bani serbet, wereng coklat
4	"RT [REDACTED] Ada CEBONG GENDHENG NGAMUK2 belain si HOMO PENISTA AGAMA!! HAAAAHA"	cebong, penista agama
5	#artikata: - cebong = anak katak yang masih berinsang - kampret = kelelawar kecil pemakan serangga #etimologi julukan: - cebong: Jokowi didukung anak katak (cebong) karena hobi memelihara katak. - kampret: Prabowo didukung partai dalam Koalisi Merah Putih/KMP (KaMPret).	-
6	[REDACTED] Mana aku tahu, aku kan cebong..	cebong
7	[REDACTED] Kaum cebong kapir udah keliatan dongoknya dari awal tambah dongok lagi hahahah	cebong
8	[REDACTED] Kampret lu zonk...dimata lu selalu salah pakde jokowi dan pemerintahan ini. Wowo lah yg benar dimata lu	-
9	[REDACTED] bani kampret bani kecebong..... pilih bani adam lah..... soale manusia	bani kampret, bani kecebong
10	[REDACTED] Wkwkwk, ke trigger jg bani onta	bani onta
11	Dapet wi-fi rumah sakit, emang saik lah 🎵 \ (・̀モ・̀) / 🎵 walupun ini rumah sakit orang-orang nya kek kampret (ಠ_ಠ) ☹ ☹	-

12	Kenapa hanya Ahok, Sukmawati dan Ganjar Pranowo (walau kecele) yang mereka sebut penista agama karena menguntungkan secara politik. Ahok dalam pilgub Jakarta yang penuh aroma rasis.	-
13	Seragamnye kotak2 kayak taplak meja	-
14	<p>██████████ Coba kalo ahok yg ngomong gtu pasti bani kampret pada murka.</p>  <p>#alumni212 #KAMPRET</p>	bani kampret, kampret
15	Fans beratnya @banitaplak pada ngamuukkkk 🤔🤔🤔🤔	-

Informasi Tambahan:

- Perhatikan bahwa kata kasar belum tentu termasuk kode kebencian (*hate code*). Kata kasar seperti **bangsat**, **anjing**, **tai**, **taik**, **bacot** (sebutin yg lumayan banyak terus dll) tidak termasuk *hate code*. Kata kasar yang termasuk *hate code* adalah kata kasar yang memang merujuk kesuatu individu atau kelompok, misal cebong, kampret, dan lain-lain. Contoh kata/frasa kasar yang lebih banyak, dapat Anda liat di lampiran panduan anotasi ini. Lebih lanjut, anda bisa juga mencari di internet atau jika ragu bisa tanyakan ke CP.
- Anda mungkin menemukan twit-twit yang mengandung kata/frasa dari bahasa daerah tertentu, silahkan cari di internet makna kata/frasa tersebut untuk lebih memahami twit. Jika ragu, silahkan tanyakan ke CP.

Lampiran: Ujaran Kasar dalam Bahasa Indonesia

Ujaran kasar merupakan suatu ujaran yang mengandung kata-kata kasar atau kotor yang disampaikan kepada lawan bicara (perorangan maupun untuk golongan tertentu) baik secara lisan maupun tulisan. Ujaran kasar disampaikan untuk memaki seseorang atau suatu golongan tertentu sebagai bentuk kemarahan, kekesalan, kekecewaan, dan sebagainya. Secara umum, kata-kata kasar yang sering digunakan oleh masyarakat Indonesia berasal dari suatu kondisi yang kurang mengenakan atau sesuatu yang menjijikan dan diharamkan oleh suatu agama (Wijana dan Rohmadi, 2011). Berikut adalah tipe-tipe kata kasar yang sering digunakan oleh masyarakat Indonesia⁴:

1. Kondisi

Kata-kata yang mengekspresikan suatu kondisi yang tidak menyenangkan biasanya digunakan sebagai kata-kata kasar oleh masyarakat Indonesia. Secara umum, beberapa kondisi kurang menyenangkan tersebut meliputi kondisi kelainan mental (contoh: *goblok*, *gila*, *bodoh*, dan lain-lain), kelainan seksual (contoh: *lesbian*, *homo*, *banci*, dan lain-lain), disabilitas (contoh: *buta*, *budek*, *bolot*, *bisu*, dan lain-lain), sifat-sifat buruk seseorang yang dilarang dan terkutuk oleh agama (contoh: *berengsek*, *bejat*, *terkutuk*, *jahanam*, *najis* dan lain-lain), serta kondisi yang berkaitan dengan peristiwa tidak menyenangkan yang dialami oleh seseorang seperti *mati*, *modar*, *mampus*, *sialan*, dan lain-lain.

2. Hewan

Dalam percakapan yang mengandung ujaran kasar, masyarakat Indonesia tidak menggunakan semua binatang sebagai referensi kata kasar. Nama hewan yang biasanya digunakan sebagai kata kasar adalah hewan-hewan yang memiliki karakteristik buruk seperti menjijikan dan diharamkan oleh suatu agama (contoh: *anjing*, *babi*), pengganggu (contoh: *bangsat*, *monyet*, *kunyuk*, dan lain-lain), parasit (contoh: *lintah*), bernaafsu tinggi (contoh: *bandot*), dan berisik (contoh: *beo*).

3. Makhluk halus

Nama-nama makhluk halus yang sering dijadikan sebagai kata-kata kasar adalah makhluk halus yang sifatnya mengganggu dan menyesatkan manusia seperti *setan*, *setan alas*, *iblis*, *tuyul*, dan lain-lain.

4. Objek tertentu

Sama seperti hewan, objek-objek yang sering dijadikan sebagai kata kasar adalah objek-objek yang bersifat menjijikan dan mengganggu kehidupan sehari-hari seperti objek yang berbau tidak sedap (contoh: *tai*, *tai kucing*, dan lain-lain), kotor dan usang (contoh: *gembel*, *gombal*), dan suara yang sangat mengganggu (contoh: *sompret*).

5. Bagian tubuh

Bagian tubuh yang sering digunakan sebagai kata-kata kasar biasanya erat kaitannya dengan kegiatan seksual seperti istilah-istilah kemaluan lelaki, istilah-istilah kemaluan wanita, dan istilah-istilah yang merujuk pada rambut kemaluan. Bagian tubuh lainnya yang sering digunakan sebagai kata kasar adalah frasa *matamu*, yang digunakan untuk memaki seseorang yang melakukan kesalahan karena tidak menggunakan penglihatannya dengan baik. Frasa lain yang sering digunakan sebagai kata kasar antara lain *hidung belang* (untuk memaki laki-laki

⁴ Untuk mengetahui arti setiap kata kasar yang dituliskan, pembaca dapat mencari di Kamus Besar Bahasa Indonesia (KBBI): <https://kbbi.kemdikbud.go.id/>

yang sering berganti-ganti pasangan) dan *mata duitan* (untuk memaki orang yang lebih mementingkan uang dalam mengerjakan sesuatu).

6. Kekerabatan

Masyarakat Indonesia biasanya menambahkan imbuhan *-mu* pada kata-kata yang merujuk kekerabatan yang lebih tua untuk memaki seseorang, seperti *ibumu*, *bapakmu*, *nenekmu*, dan *kakekmu*.

7. Aktivitas

Jenis-jenis aktivitas yang sering digunakan sebagai kata kasar oleh masyarakat Indonesia adalah istilah-istilah yang merujuk pada aktivitas seks, seperti istilah-istilah yang maknanya merujuk ke aktivitas senggama dan masturbasi.

8. Profesi

Profesi-profesi yang diharamkan oleh agama biasanya dijadikan sebagai kata kasar oleh sebagian masyarakat Indonesia, seperti *maling*, *sundal*, *bajingan*, *pelacur*, *jablay*, dan lain-lain.