

Periode : Semester Ganjil / Genap  
Tahun : 2020  
Skema Penelitian : Hibah Internal  
Tema RIP Penelitian : Big Data (Machine Learning)

**LAPORAN PENELITIAN MANDIRI**

**SISTEM DETEKSI DINI EMOTIONAL DISTRESS DI MEDIA SOSIAL TWITTER**



**PENGUSUL**

**Ir. Munawar., MMSI., M.Com., PhD (0324066901)**

**Fakultas Ilmu Komputer  
Teknik Informatika  
Universitas Esa Unggul  
2020**

## HALAMAN PENGESAHAN

- 
1. Judul Penelitian : Sistem Deteksi Dini *Emotional Distress* di Media Sosial Twitter
- 
1. Ketua Peneliti
- a. Nama lengkap dengan gelar : Ir. Munawar MMSI., M.Com., PhD
  - b. Pangkat/Gol/NIP :
  - c. Jabatan Fungsional/Struktural : Lektor Kepala
  - d. Program Studi/Jurusan : Informatika
  - e. Fakultas : Fasilkom
  - f. Alamat Rumah/HP : 08128100435
  - g. E-mail : [an\\_moenawar@yahoo.com](mailto:an_moenawar@yahoo.com)
- Anggota Peneliti
- a. Nama lengkap dengan gelar : -
  - c. Jabatan Fungsional/Struktural : -
  - d. Program Studi/Jurusan :
- 
3. Jumlah Tim Peneliti : 1 orang
- 
4. Lokasi Penelitian : Jakarta dan sekitarnya
- 
5. Kerjasama (kalau ada)
- a. Nama Instansi : -
  - b. Alamat : -
- 
6. Jangka waktu penelitian : 12. bulan
- 
7. Biaya Penelitian : Rp. 14.500.000,00 (Empat Belas Juta Lima Ratus Ribu Rupiah)
- 

Mengetahui,  
Dekan Fakultas Ilmu Komputer



Habibullah Akbar S.Si, MSc, PhD  
NIK: 218030726

Jakarta, 24 Agustus 2020

Peneliti,



Ir. Munawar, MMSI., M.Com., PhD  
NIK: 202080208

Menyetujui,  
Ketua Lembaga Penelitian dan Pengabdian kepada Masyarakat  
Universitas Esa Unggul



Dr Erry Yudhya Mulyani, MSc  
NIK. 209100388

## DAFTAR ISI

	Halaman
Judul.....	1
Halaman Pengesahan.....	2
Daftar Isi.....	3
Abstrak.....	4
<b>Bab 1. Pendahuluan.....</b>	<b>5</b>
1.1. Latar Belakang.....	5
1.2. Permasalahan.....	5
1.3. Tujuan Penelitian.....	6
1.4. Ruang Lingkup.....	6
1.5. Kebaruan dan Terobosan Teknologi.....	6
1.6. Manfaat Penelitian.....	7
1.7. Luaran Penelitian.....	7
<b>Bab 2. Renstra dan Peta Jalan Penelitian Fakultas Ilmu Komputer.....</b>	<b>8</b>
2.1. Renstra Fakultas Ilmu Komputer.....	8
2.2. Peta Rencana Penelitian.....	10
<b>Bab 3. Tinjauan Pustaka.....</b>	<b>11</b>
3.1. Pengertian <i>Emotioanl Distress</i> .....	11
3.2. Ekspresi di media sosial sebagai bentuk ekspresi diri.....	11
3.3. Kamus <i>Emotional distress</i> .....	12
3.4. DASS-21.....	12
3.5. Penelitian Sebelumnya dan Perbedaannya dengan Usulan.....	13
<b>Bab 4. Metode Penelitian.....</b>	<b>14</b>
<b>Bab 5. Hasil dan Pembahasan Aplikasi Crawler.....</b>	<b>17</b>
5.1. Algoritma Aplikasi Crawler.....	17
5.2. Hasil Pengujian Aplikasi Crawler.....	19
5.3. Pembahasan.....	22
<b>Bab 6. Kesimpulan dan Saran.....</b>	<b>24</b>
6.1. Kesimpulan.....	24
6.2. Saran.....	24

Daftar Pustaka.....	25
Lampiran 1. Quesiner DASS-21.....	27

## Sistem Deteksi Dini *Emosional Distres* Di Media Sosial Twitter

### Abstrak

Ekspresi seseorang di media sosial memberikan isyarat psikologis penting tentang status kesehatan mental mereka. Dengan besarnya pengguna media sosial di Indonesia yang mencapai 150 juta (56% dari total populasi) bisa digunakan sebagai indikator untuk identifikasi *emotional distress* yang ada di masyarakat. Penelitian ini dimaksudkan untuk menarik data dari postingan di media sosial Twitter tanpa peran aktif responden untuk mengisi kuesioner atau berpartisipasi aktif dalam wawancara profesional guna mendapatkan tingkatan *emotional distress* masyarakat seperti depresi, kecemasan dan stres.

Laporan penelitian ini difokuskan kepada pembuatan mesin *crawling* data dari media sosial Twitter sebagai *milestone* pertama dari penelitian ini. Dari hasil mesin *crawling* ini dapat digunakan untuk menarik cuitan pengguna Twitter berdasarkan kata kunci tertentu. Kata kunci yang digunakan didasarkan kepada ekstraksi kuesioner DASS-21 (*Depression Anxiety Stress Scales*) sebagai ungkapan *emotional distress* yang ada di media sosial Twitter.

Secara umum mesin *crawling* ini menggunakan bahasa pemrograman Python dengan memanfaatkan library Tweepy untuk menarik data cuitan dari Twitter. Hasil yang diperoleh adalah file dengan format JSON. Hasil dari *crawling* ini selanjutnya dilakukan pemrosesan yang meliputi: *cleansing* (pembersihan tanda baca), *case folding* (mengubah menjadi huruf kecil), *tokenize* (pemotongan kalimat menjadi kata), *stop word* (penghapusan kata penghubung dan kata tidak bermakna), *stemming* (pengambilan kata dasar). Hasil akhirnya kemudian divisualisasikan ke dalam grafik untuk melihat trend dan kecenderungan dari penggunaan kata kunci tadi pada satu periode waktu tertentu. Hasil ini akan diproses di *milestone* penelitian berikutnya untuk melihat trend *emotional distress* di media sosial Twitter.

Kata kunci : *crawling*, *emotional distress* , DASS-21, media sosial.

## BAB I. PENDAHULUAN

### 1.1. Latar Belakang

Berdasarkan hasil riset Wearesosial Hootsuite yang dirilis Januari 2019 pengguna media sosial di Indonesia mencapai 150 juta atau sebesar 56% dari total populasi. Jumlah tersebut naik 20% dari survei sebelumnya. Sementara pengguna media sosial mobile (gadget) mencapai 130 juta atau sekitar 48% dari populasi (<https://wearesocial.com/global-digital-report-2019>).

Disisi lain, kata-kata yang digunakan orang di media social memberikan isyarat psikologis penting untuk status kesehatan mental mereka (Jarrold et al., 2011). Banyak penelitian menunjukkan adanya hubungan *emotional distress* dan kecenderungan bunuh diri dengan pola linguistik dalam posting media social (Sueki Hajime, 2015). Dengan demikian fitur linguistik yang diperoleh dari media sosial dapat digunakan sebagai indikator untuk mengidentifikasi *emotional distress* yang ada di masyarakat.

Dalam *Global Health Data Exchange 2017*, prevalensi gangguan mental di Indonesia pada usia produktif (20-54 tahun) mencapai 15,91 juta orang atau 58,3 persen dari total penderita gangguan mental semua umur (27,26 juta orang). Prevalensi gangguan jiwa juga terekam dalam Riset Kesehatan Dasar (Riskedas) oleh Kementerian Kesehatan tahun 2018. Dalam Riskedas 2018 itu prevalensi gangguan mental emosional yang ditunjukkan pada usia 15 tahun ke atas mencapai 9,8 persen dari jumlah penduduk. Angka ini meningkat dibandingkan dengan tahun 2013, yaitu sebesar 6 persen. Adapun prevalensi depresi pada penduduk untuk umur 15 ke atas mencapai 6,1 persen dari jumlah penduduk dan prevalensi gangguan jiwa berat, seperti skizofrenia, mencapai 7 per 1.000 penduduk

Untuk menilai risiko bunuh diri dan emosional distres, banyak alat bantu yang telah dikembangkan dan divalidasi (Cheng, Li, Kwok, Zhu, & Yip, 2017). Namun sayangnya, alat-bantu ini sering membutuhkan responden untuk mengisi kuesioner atau berpartisipasi dalam sebuah wawancara profesional. Padahal orang mengalami *emotional distress* atau memiliki kecenderungan bunuh diri sering memiliki motivasi yang rendah untuk mencari bantuan dari para profesional (Rickwood, Deane, & Wilson, 2007).

### 1.2. Permasalahan

Penggunaan media sosial yang kian meluas telah memberikana ruang gerak yang alami bagi seseorang untuk mengekspresikan luapan emosinya sehingga bisa dijadikan sebagai acuan untuk mempelajari berbagai masalah kesehatan mental dan kecenderungan bunuh diri

(Cheng, Chang, & Yip, 2012). Seseorang yang tertekan kondisi kejiwaannya cenderung untuk memberitahukan kepada orang lain tentang pemikiran mereka untuk bunuh diri (Cheng, Kwok, Zhu, Guan, & Yip, 2015). Hal ini sejalan dengan studi psiko-linguistik yang melihat bahwa kata-kata atau bahasa bisa dijadikan sebagai tanda yang bermakna dalam menyampaikan apa yang ada dalam pemikiran kita (Pennebaker, Mehl, & Niederhoffer, 2003)

Kasus-kasus penggunaan media social untuk mengekspresikan social distress, cara bunuh diri bahkan siaran langsung bunuh diri sudah pernah dilaporkan dan dipelajari di berbagai negara (Sueki Hajime, 2015). Beberapa studi sebelumnya menunjukkan adanya penggunaan media sosial untuk menilai depresi atau resiko bunuh diri dalam versi Bahasa Inggris (Sueki Hajime, 2015); (Braithwaite, Giraud-Carrier, West, Barnes, & Hanson, 2016); (O'Dea et al., 2015). Hingga saat ini, sepanjang pengetahuan penulis, masih belum ada penelitian tentang sistem deteksi *emotional distress* dalam bentuk aplikasi terintegrasi khususnya untuk ekspresi kata-kata dalam Bahasa Indonesia. Oleh karena itu dirasa perlu untuk membuat alat bantu yang bisa mendeteksi tekanan emosional seseorang secara alami guna mengidentifikasi secara dini adanya gangguan emosional disorder maupun kecenderungan bunuh diri yang diekspresikan di media social.

### **1.3. Tujuan Penelitian**

Penelitian ini bertujuan untuk membuat aplikasi mesin pembelajaran (*machine learning*) yang bisa digunakan secara otomatis dalam mendeteksi *emotional distress* yang diekspresikan ke media sosial Twitter khususnya yang berbahasa Indonesia.

### **1.4. Ruang Lingkup**

Agar tidak terlalu melebar, penelitian ini dilakukan di media sosial Twitter. Media sosial Twitter adalah media sosial terbanyak ketiga yang masih memungkinkan diakses datanya di Indonesia (Hotsuite, 2019).

### **1.5. Kebaruan dan Terobosan Teknologi**

Kebaharuan atau terobosan teknologi yang akan dihasilkan dalam penelitian ini diantaranya adalah:

- a) Adanya kamus khusus tentang *emotional distress* dalam bahasa Indonesia
- b) Adanya mesin crawler berbasis kumpulan kata kunci yang terkait dengan *emotional distress* dan penyakit kejiwaan yang lainnya dari berbagai media social Twitter.

- c) Adanya mesin crawler untuk menarik berita-berita yang terkait dengan berbagai tindak kejahatan di media online berbahasa Indonesia
- d) Digital dashboard untuk pemetaan *emotional distress* berdasarkan umur, gender, lokasi dan berbagai aspek demografi lainnya
- e) Digital dashboard untuk pemetaan *emotional distress* di media social Twitter versus tindak kejahatan di berbagai media online berbahasa Indonesia.

### **1.6. Manfaat Penelitian**

Penelitian ini akan memberikan dampak yang luar biasa terutama kepada para psikolog dalam memberikan penilaian atas level depresi, kecemasan dan stress pasiennya tanpa harus melalui serangkaian wawancara professional maupun pengisian questioner. Dengan demikian kondisi alami para pasiennya yang diekspresikan di media sosial diharapkan bisa merefleksikan kondisi *emotional distress* mereka.

Dampak ikutan penting lainnya adalah bisa dipetakannya *emotional distress* masyarakat yang diekspresikan di media sosial berdasarkan umur, gender, lokasi dan berbagai aspek demografi lainnya. Dengan demikian bisa dilakukan tindakan preventif terkait dengan kondisi yang ada di masyarakat.

Hal lain yang tidak kalah penting adalah pemetaan *emotional distress* di media sosial versus tindak kejahatan di berbagai media online berbahasa Indonesia. Pemetaan ini diharapkan bisa memberikan gambaran yang lebih riil tentang kondisi yang ada di masyarakat.

### **1.7. Luaran Penelitian**

Luaran (output) yang dihasilkan dari tahap pertama ini adalah mesin *crawling* untuk penarikan data cuitan pengguna di Twitter berdasarkan kata kunci tertentu yang diekstraksi dari kuesioner DASS-21 sebagai ungkapan *emotional distress*.

## BAB II. RENSTRA DAN PETA JALAN PENELITIAN FAKULTAS ILMU KOMPUTER

### 2.1. Renstra Fakultas Ilmu Komputer

Roadmap Penelitian dan PkM Fasilkom-UEU meliputi topik-topik yang sesuai dengan bidang ilmu yang telah dijelaskan diatas. Semua kegiatan penelitian dan PkM di Fasilkom-UEU adalah mengacu kepada topik yang memayunginya di tingkat kelompok studi. Kegiatan pengabdian kepada masyarakat merupakan implementasi hasil kegiatan penelitian. Kegiatan penelitian dan PkM merupakan satu paket yang tidak terpisahkan.

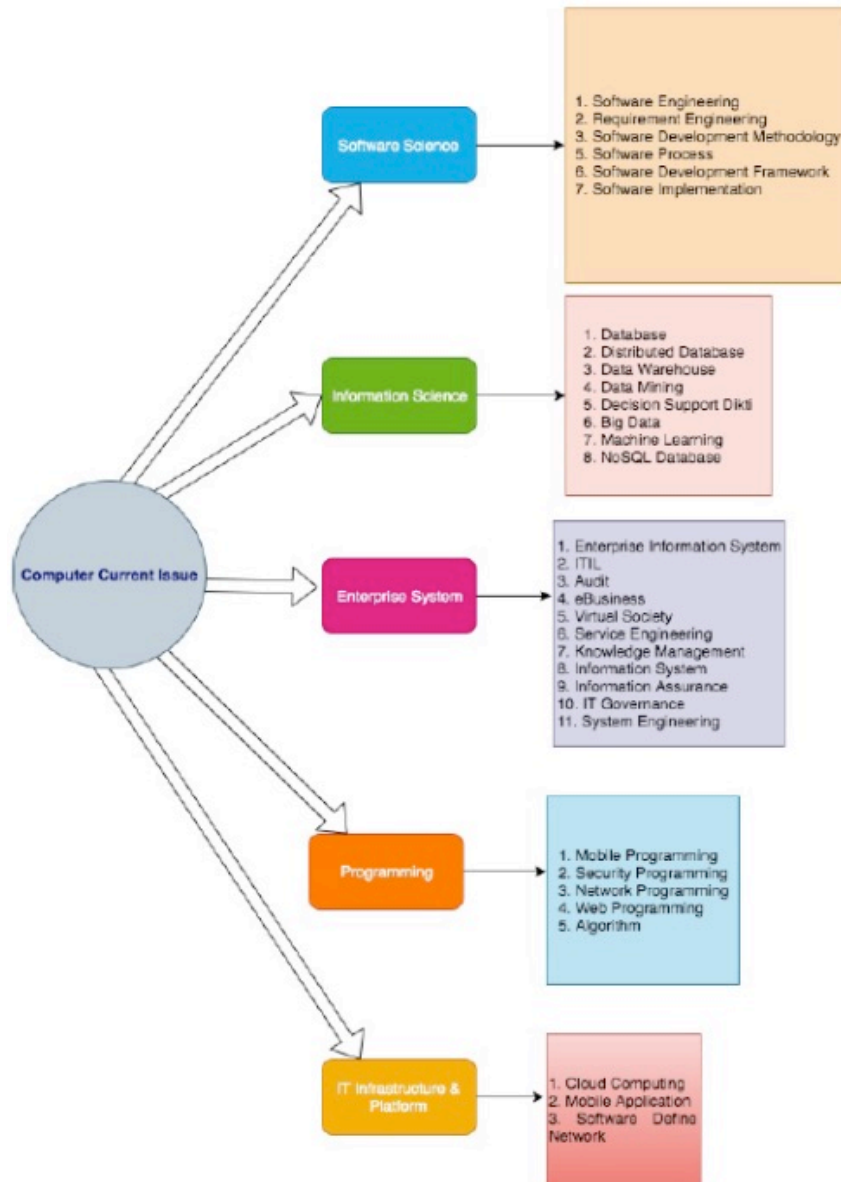
Rencana kegiatan penelitian di setiap kelompok studi didasarkan kepada beberapa kompetensi yang berbeda. Beberapa kelompok studi yang ada di Fasilkom-UEU mengadakan penelitian berbasis laboratorium, beberapa yang lain berbasis kelompok dosen. Payung Penelitian dan Pengabdian kepada Masyarakat di Tingkat Fasilkom-UEU diuraikan seperti Nampak di Gambar 2.1. Sedangkan rencana kerja dan tahapan penelitian Fakultas Ilmu Komputer tersaji di Tabel 2.1.

Tabel 2.1. Rencana kerja dan tahapan penelitian Fakultas Ilmu Komputer

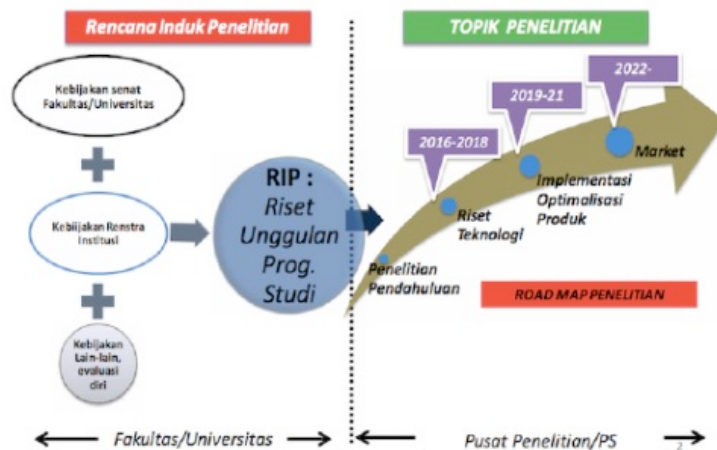
Road Map Penelitian								
Kluster	Tema	2017	2018	2019	2020	2021	Pelaksana	
Industri dan Masyarakat								
Technopreneurship								
Technopreneurship	Penelitian Desentralisasi	Software Science	Software Engineering, Requirement Engineering	Software Development Methodology	Software Process	Software Development Framework	Software Implementation	Fasilkom-UEU Semua kelompok, bidang studi, laboratorium dalam kegiatan mandiri maupun terpadu
	Penelitian Kompetitif Nasional	Information Science	Database, Distributed Database	Data Warehouse, Data Mining	Decision Support System, Big Data	Machine Learning	No SQL Database	
	Insinas	Enterprise System	Enterprise Information System	ITIL, Audit, eBusiness, Virtual Society	Service Engineering, Knowledge Management	Information System, Information Assurance	IT Governance, System Engineering	
	Pengabdian kepada Masyarakat	Programming	Mobile Programming	Security Programming	Network Programming	Web Programming	Algorithm	
	Penelitian Internal	IT Infrastructure & Platform	Cloud Computing	Mobile Application	Software Define Network			
Dosen Aktif dan Mahasiswa Fasilkom-UEU								

Dengan melihat konsentrasi penelitian ini kepada big data dan *machine learning*, terlihat bahwa penelitian ini sejalan dengan road map, rencana kerja dan tahapan penelitian Fakultas Ilmu Komputer.





Gambar 2.1. Road Map Penelitian Fakultas Ilmu Komputer



Gambar 2.2. Hubungan antara RIP dengan roadmap penelitian

## 2.2 Peta Rencana Penelitian

Penelitian ini direncanakan akan berlangsung selama 2 tahun, dimana pada tahun pertama akan difokuskan untuk membuat mesin crawling data Twitter berdasarkan kata kunci pada ekstraksi DASS-21. Pada tahun kedua difokuskan kepada pengumpulan data cuitan di Twitter untuk dipetakan berdasarkan kuesioner DASS-21 guna mendapatkan trend dan peta emotional distress seperti depresi, kecemasan, dan stress dari postingan masyarakat yang ada di media sosial Twitter sehingga bisa didapatkan petanya sesuai dengan kondisi demografi. Peta penelitian ini secara lebih terinci ditunjukkan pada Tabel 2.2.

Tabel 2.2. Road Map Usulan Penelitian

<b>Tahun 1</b> TKT level 1–3	<b>Tahun 2</b> TKT level4-6
Pembuatan mesin crawling untuk penarikan postingan dari cuitan di Twitter dengan kata kunci yang di ekstrak dari quesioner DASS-21	Pembuatan aplikasi machine learning untuk pemetaan depresi, kecemasan, dan stress sesuai dengan kondisi demografi berbasis pada ekspresi yang diposting di media sosial
Program Utama: <ul style="list-style-type: none"><li>• Pembuatan aplikasi crawler untuk Twitter</li><li>• Pembuatan aplikasi pengolahan data hasil crawling (cleansing, tokenizing, stopword dan stemming).</li></ul>	Program Utama: <ul style="list-style-type: none"><li>• Pembuatan aplikasi machine learning dengan memanfaatkan NLP (Neuro Language Program), sentiment analysis dan SVM classifier.</li><li>• Pembuatan dashboard pemetaan kondisi depresi, kecemasan, dan stress dari hasil postingan sosial media Twitter</li><li>• Mengintegrasikan kamus <i>emotional distress</i> , mesin crawler dan machine learning serta digital dashboard sebagai satu kesatuan aplikasi berbasis web sehingga bisa diakses dari mana saja dan kapan saja.</li></ul>

### BAB III. TINJAUAN PUSTAKA

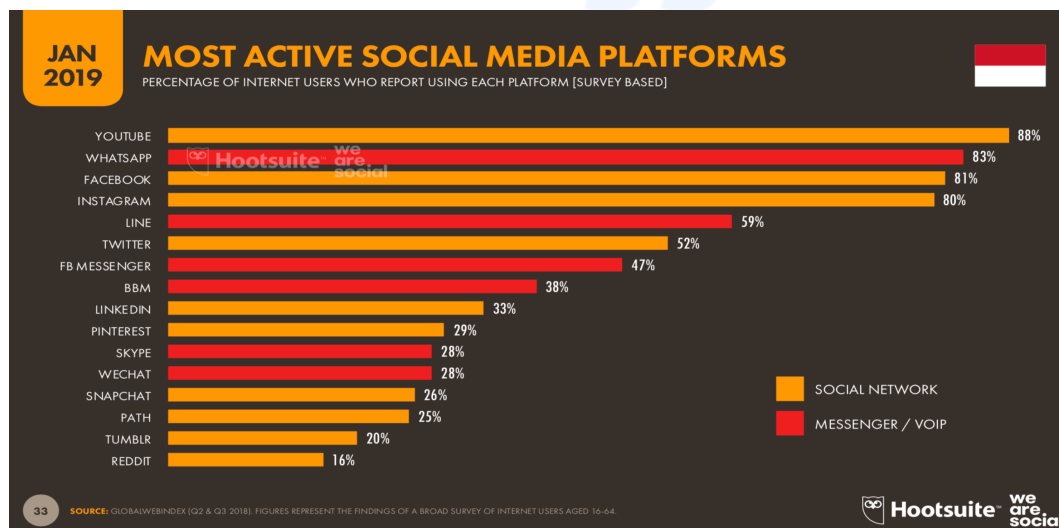
#### 3.1. Pengertian *Emotional Distress*.

*Emotional distress* merupakan reaksi emosional individu ketika menghadapi tekanan-tekanan baik yang bersifat internal maupun eksternal. (R.S, 1993) menyatakan *emotional distress* sebagai stres psikologis yang merupakan reaksi terhadap berbagai jenis ancaman yang muncul dari dalam diri dan lingkungan. Reaksi tersebut terjadi akibat evaluasi terhadap sesuatu yang dianggap mengancam kesejahteraan (*wellbeing*) individu. (Mirowsky, J., Ross, 2003) dan (McCraty, 2006) mendefinisikan *emotional distress* sebagai keadaan ketika seseorang mengalami emosi negatif sebagai respon atas stres yang dialami. *Emotional distress* sebagai trauma mental atau psikologis yang disebabkan oleh perilaku yang menyakitkan (*tortious*) atau tidak menyakitkan (*non-tortious*) (Bogoroch, 2005)

#### 3.2. Ekspresi di media sosial sebagai bentuk ekspresi diri

Media sosial saat ini sudah menjadi media alternative untuk mengekspresikan diri atau berbagi informasi dengan yang lain. Dengan penggunaan media social yang kian masif, ekspresi diri di media sosial bisa menjadi indikasi status kesehatan mental mereka (Jarrold et al., 2011), oleh karena itu bisa dijadikan sebagai acuan untuk mempelajari berbagai masalah kesehatan mental dan kecenderungan bunuh diri (Cheng et al., 2012). Hal ini sejalan dengan teori psiko-linguistik yang menyatakan bahwa kata-kata atau bahasa merefleksikan apa yang ada dalam pemikiran kita (Pennebaker et al., 2003).

Media sosial apa saja yang dimanfaatkan oleh mayoritas masyarakat Indonesia bisa dilihat pada Gambar 3.1.



Gambar 3.1. Media Sosial yang diakses di Indonesia (sumber: Hootsuite, 2019)

### 3.3. Kamus *Emotional distress*

Pembuatan kamus untuk tujuan tertentu secara umum sangat berguna dalam identifikasi kasus seperti halnya kamus bunuh diri di Cina ((Lv, Li, Liu, & Zhu, 2015) untuk mendeteksi resiko bunuh diri di media social Weibo, penggunaan LIWC (Linguistic Inquiry and Word Count) untuk penggunaan umum (Cheng et al., 2017). Untuk analisis atas bahasa yang diekspresikan di sosial media, telah ada alat bantu khusus yang disebut dengan NLTK (Natural Language ToolKit) dalam banyak Bahasa, termasuk dalam Bahasa Indonesia. Namun sepanjang pengetahuan penulis, hingga saat ini belum ada kamus khusus dalam Bahasa Indonesia tentang *emotional distress* .Dirasa perlu dibuat kamus khusus tentang *emotional distress* untuk menganalisis ekspresi yang diungkapkan di media social.

Untuk menilai *emotional distress* di media social, saat ini sudah cukup banyak diantaranya questioner ide bunuh diri, skala kemungkinan bunuh diri, skala depresi dan kekhawatiran dan lain-lain (Cheng et al., 2017). Namun sayang, semua teknik ini membutuhkan partisipasi aktif responden untuk mengisi questioner atau berpartisipasi aktif dalam wawancara professional. Padahal orang-orang yang stress atau memiliki gangguan kejiwaan seringkali memiliki motivasi yang rendah untuk mencari bantuan kepada professional(Rickwood et al., 2007). Dibutuhkan suatu alat bantu yang bisa menarik kalimat-kalimat yang diposting di berbagai media sosial (*crawler*) untuk dinilai tingkat resiko *emotional distress* nya, bahkan jika perlu bisa dikhususkan penarikan kalimat dikhususkan untuk pengguna tertentu dalam kondisinya yang alami tanpa perlu harus konsultasi maupun mengisi questioner guna pencegahan dari hal-hal yang kurang diinginkan setelah ditengarai sudah mencapai level tertentu dalam *emotional distress* .

### 3.4. DASS-21

DASS-21 adalah bentuk pendek set dari 3 skala *self-report* yang dirancang untuk mengukur keadaan emosional dari depresi, kecemasan, dan stress (Henry & Crawford, 2005). DASS-21 dibuat bukan hanya sebagai skala biasa untuk mengukur kondisi emosional secara konvensional, tetapi juga lebih jauh sebagai proses untuk mengidentifikasi, mengerti, dan mengukur keadaan emosional secara klinis yang sedang dialami, yang biasanya disebut sebagai depresi, kecemasan, dan stress. Hanya saja, DASS-21 ini membutuhkan peran aktif responden karena wujudnya adalah questioner. Untuk menjembatani hal tersebut, akan diambil kata kunci utama per masing-masing pertanyaan. Dari kata kunci tersebut selanjutnya digunakan untuk menarik semua postingan yang ada di media sosial selama 4-5 bulan. Hasil *crawling* data akan

dibobotkan sesuai ketentuan DASS-21 sehingga bisa didapatkan peta kondisi depresi, kecemasan, dan stress sesuai dengan kondisi demografi.

### 3.5. Penelitian Sebelumnya dan Perbedaannya dengan Usulan

Beberapa penelitian sebelumnya menunjukkan adanya potensi penggunaan media sosial untuk menilai resiko depresi dan bahkan resiko bunuh diri dalam Bahasa Inggris (Sueki Hajime, 2015)(Braithwaite et al., 2016)(Gamon & Counts, 2013). Hanya sayang, hingga saat ini belum ada untuk yang berbahasa Indonesia.

Penggunaan kamus untuk mendeteksi resiko bunuh diri di media sosial dalam Bahasa Cina juga sudah dilakukan (Lv et al., 2015). Hanya sayang, validasi akhir kamus tersebut dilakukan oleh mahasiswa pasca sarjana yang belum terlatih dalam pencegahan bunuh diri. Butuh pengujian secara empiris untuk bisa menggunakan kamus tersebut dalam kasus pencegahan bunuh diri.

Pendekatan lain untuk penilaian resiko bunuh diri dan *emotional distress* adalah menggunakan *text mining* dan *machine learning*(Cheng et al., 2017). Mereka menggunakan pendekatan survey kepada pengguna Weibo di Cina untuk menilai resiko bunuh diri dan *emotional distress*. Kenyataannya, jangankan mau ikut dalam survey, orang-orang yang stress seringkali memiliki motivasi yang rendah untuk mencari bantuan kepada professional (Rickwood et al., 2007). Disisi lain, besarnya jumlah pengguna media sosial bisa jadi juga merupakan orang yang beresiko terkena *emotional distress* atau bahkan beresiko bunuh diri (Lv et al., 2015). Dengan sifat media sosial yang bisa diakses setiap saat, informatif dan adanya kecenderungan sebagian penderita gangguan emosional untuk menggunakan media sosial sebagai media untuk mengekspresikan emosi mereka termasuk rencana dan pikiran-pikiran bunuh diri (G, 2014) maka pengambilan data di media sosial secara otomatis (*crawling*) tanpa keterlibatan *suspect emotional distress*, diharapkan bisa memetakan percakapan/ posting yang ada di media sosial khususnya yang terkait dengan *emotional distress* secara alami.

## BAB IV. METODE PENELITIAN

Metode penelitian yang digunakan didasarkan kepada masalah dan tujuan yang ingin dicapai. Untuk tahun pertama difokuskan untuk pembangunan aplikasi *crawler* (penarikan data dari medis sosial Twitter secara otomatis berbasis kata kunci) sedangkan di tahun kedua adalah pembuatan aplikasi machine learning untuk klasifikasi dan pemetaan kondisi depresi, kecemasan, dan stress dari hasil postingan sosial media Twitter.

Secara umum persoalan utama di tahun pertama adalah rendahnya motivasi para penderita *emotional distress* untuk mencari bantuan kepada professional. Disisi lain, metode yang biasa digunakan para professional untuk mengetahui level *emotional distress* penderita adalah pengisian questioner yang juga di jauhi para penderita. Untuk itu questioner DASS-21 dimodifikasi sedemikian rupa sehingga tidak berbentuk questioner lagi namun menjadi kumpulan kata kunci yang bisa mencerminkan semua elemen DASS-21 agar bisa digunakan sebagai kata kunci penarikan data postingan dari Twitter. Secara lebih lengkap, metode penelitian yang digunakan bisa dilihat pada Gambar 4.1. Adapun penjelasan lebih rinci tentang metode penelitian ini adalah sebagai berikut:

a. Ekstraksi kata kunci dari DASS-21

Dengan bantuan psikolog, kuesioner DASS-21 dieksplorasi lebih dalam guna mendapatkan kata kunci utama dan frekuensinya sehingga didapatkanlah kumpulan kata kunci dan frekuensi nya

b. Crawling data Twitter

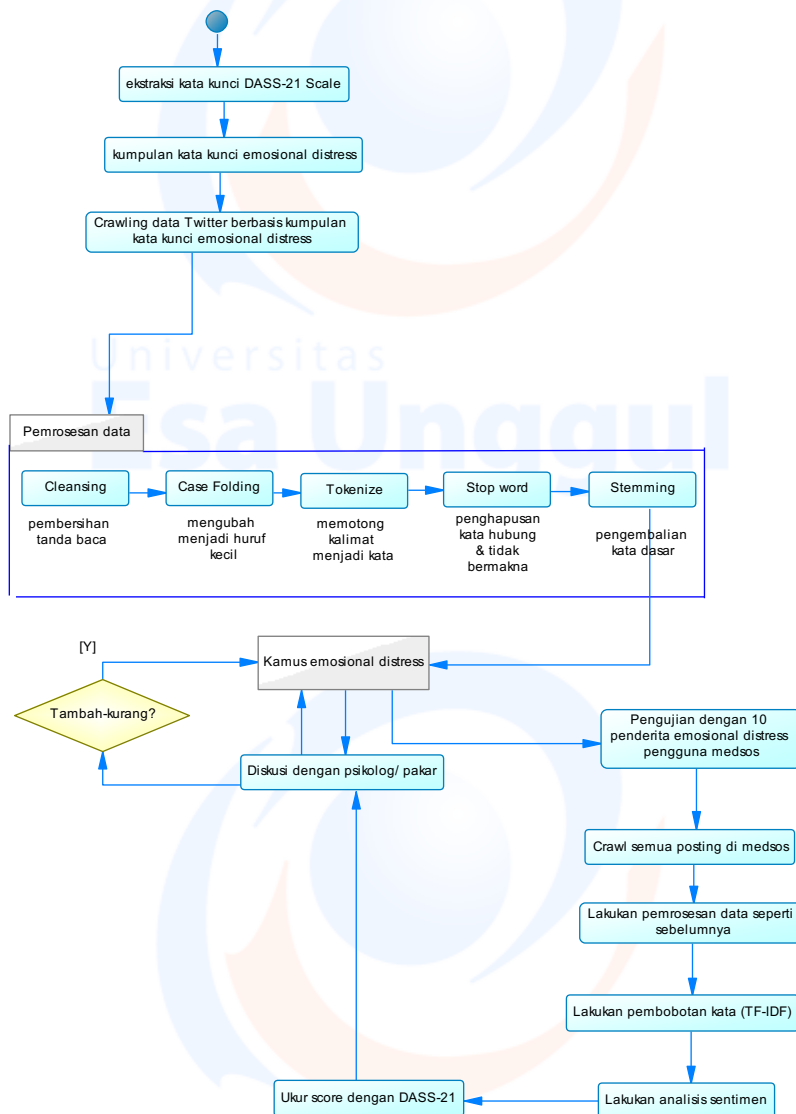
Atas dasar kata kunci dari tahap a di atas selanjutnya dilakukanlah crawling data Twitter selama beberapa periode. Periode pengumpulan data akan dilakukan selama 3 bulan berturut-turut di tahun 2020.

c. Pemrosesan Data

Data yang terkumpul dari tahap b selanjutnya akan dilakukan pemrosesan data dengan cara sebagai berikut:

- Cleaning : proses ini dilakukan dengan cara pembersihan dari tanda baca yang tidak perlu
- Case Folding : proses ini mengubah semua huruf dari data yang terkumpul menjadi huruf kecil semua
- Tokenize : proses ini akan memotong kalimat tweet menjadi kata-kata
- Stop word : proses ini akan menghapus kata hubung dan kata-kata tidak bermakna

- Stemming : proses ini akan mengembalikan hasil pada tahap sebelumnya menjadi kata dasar
- d. Pembuatan kamus *emotional distress*
- Kamus dasar hasil ekstraksi dari cuitan (*tweet*) netizen di Twitter selanjutnya didiskusikan dengan pakar psikologi untuk justifikasi akhir sehingga bisa didapatkan kamus *emotional distress* versi awal
- e. Pengujian kamus *emotional distress* dengan kasus nyata
- Hasil kamus *emotional distress* awal selanjutnya akan diujikan kepada 10 orang penderita *emotional distress*. 10 orang ini adalah rekomendasi psikolog klinis yang benar-benar mengalami *emotional distress* sekaligus juga pengguna media sosial Twitter. Keseluruhan cuitan mereka kemudian ditarik dan dilakukan pemrosesan data seperti pada tahap c di atas. Perbedaannya dengan tahap c adalah dilakukannya pembobotan kata (TF-IDF) dan dilakukan analisis sentiment untuk mengetahui apakah cuitan penderita *emotional distress* ini positif, negative atau netral. Hasilnya kemudian diukur dengan score DASS-21 untuk menunjukkan derajat *emotional distress* mereka. Sebagai finalisasi dilakukan justifikasi hasil ini kepada psikolog klinis untuk melakukan tambah kurang atas kamus *emotional distress* yang sudah terbentuk.



Gambar 4.1. Metode Penelitian



## BAB V. HASIL DAN PEMBAHASAN APLIKASI CRAWLER

### 5.1. Algoritma Aplikasi Crawler

Secara umum aplikasi crawler yang dibangun didasarkan kepada tahapan-tahapan berikut ini

#### 1. Proses crawling

Aplikasi ini dibangun dengan Bahasa pemrograman Python dengan memanfaatkan library Tweepy yang mengakses micro blogging Twitter melalui Twitter API. Data yang diperoleh berupa file dengan format JSON yang didalamnya terkandung *id, text, source, username, user location, user description, user time-zone, followers*.

Proses pengumpulan datanya didasarkan kepada kata kunci yang diekstraksi dari kuesioner DASS-21 untuk mendapatkan data depresi, kecemasan, dan stress sesuai dengan kondisi demografi dari hasil cuitan di Twitter.

#### 2. Data pre-processing

Data hasil pengumpulan dari aplikasi crawler selanjutnya dibersihkan agar sesuai dengan kebutuhan penelitian agar mudah untuk dianalisis. Berikut tahapannya:

- *Cleaning*

*Cleaning* bertujuan untuk menghilangkan tanda baca, username (@), url, dan karakter tertentu (emoticon) dari *tweets* yang sudah dikumpulkan. Pembuatan code untuk mengkonfigurasi sesuatu yang ingin dihapus ini bisa menggunakan kaedah penulisan Regular Expression (RegEx). Penggunaan *library re* dan *string* bisa digunakan untuk mempermudah pengenalan karakter dari isi dokumen tersebut.

- *Case Folding*

Dalam text preprocessing proses *case folding* bertujuan untuk mengubah semua huruf dalam sebuah dokumen teks menjadi huruf kecil (*lowercase*) dengan menggunakan fungsi *lower()* pada python.

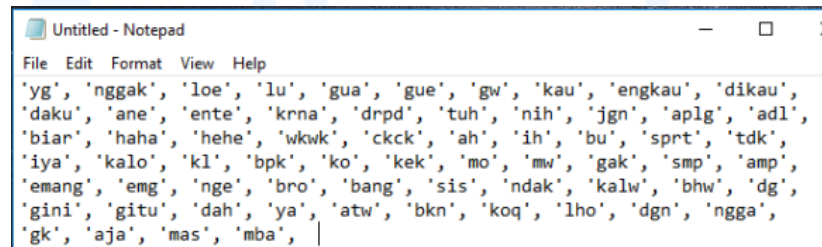
- Tokenize

Dokumen teks yang terdiri dari sekumpulan kalimat perlu dilakukan proses *tokenization* guna memecah dokumen tersebut menjadi bagian-bagian kata yang disebut token. Proses tokenization bisa dilakukan dengan bantuan *library NLTK* (Natural Language Toolkit).

- Stopword

Tahapan ini adalah pengambilan kata-kata yang dianggap penting atau membuang kata-kata yang dianggap tidak terlalu mempunyai arti penting dalam proses *text mining*. *Stopword Removal* pada penelitian ini dilakukan dengan panduan kamus *stopword Indonesian* dan *English* yang di *import* dari *library NLTK*.

Karena postingan di Twitter banyak menggunakan Bahasa gaul, maka perlu ditambahkan beberapa kata dalam stop word. Berikut ini adalah tambahan kata yang digunakan



```
Untitled - Notepad
File Edit Format View Help
'yg', 'nggak', 'loe', 'lu', 'gua', 'gue', 'gw', 'kau', 'engkau', 'dikau',
'daku', 'ane', 'ente', 'krna', 'drpd', 'tuh', 'nih', 'jgn', 'aplg', 'adl',
'bian', 'haha', 'hehe', 'wkwk', 'ckck', 'ah', 'ih', 'bu', 'sprt', 'tdk',
'iya', 'kalo', 'kl', 'bpk', 'ko', 'kek', 'mo', 'mw', 'gak', 'smp', 'amp',
'emang', 'emg', 'nge', 'bro', 'bang', 'sis', 'ndak', 'kalw', 'bhw', 'dg',
'gini', 'gitu', 'dah', 'ya', 'atw', 'bkn', 'koq', 'lho', 'dgn', 'ngga',
'gk', 'aja', 'mas', 'mba', |
```

Gambar 5.1. Penambahan kata dalam kamus stopwords

- *Stemming*

*Stemming* bertujuan untuk mentransformasikan kata menjadi kata dasarnya (*root word*) dengan menghilangkan semua imbuhan kata. Saat ini, proses *stemming* Bahasa Indonesia menggunakan NLTK belum ada, sehingga proses *stemming* penelitian ini menggunakan *library* Sastrawi.

Hasil yang didapatkan adalah struktur data penelitian dalam bentuk file *.csv* yang berisi kolom teks hasil praproses, kolom *SN (Screen Name)* yaitu akun yang membuat *tweet*, kolom *Reply to SN* yang menunjukkan kepada siapa *tweet* tersebut ditujukan serta kolom *created* yang gunanya untuk menunjukkan waktu *tweet* tersebut dipublikasikan.

### 3. Visualisasi grafik hasil crawling

Untuk visualisasi data dengan grafik dilakukan dengan cara memanfaatkan *library-library* yang ada pada python. *Library* yang digunakan sebagai berikut:

- *Library Pandas*

*Pandas* digunakan untuk membaca *file* dalam *format .csv*. *Pandas* adalah *spreadsheet* namun tidak memiliki GUI dan punya fitur seperti *SQL* yang bisa digunakan untuk pengolahan data.

- *Library Matplotlib*

Matplotlib library biasa digunakan dalam disiplin *data science* untuk menyajikan data dalam bentuk visual seperti *graph* atau *chart*. Beberapa plot digunakan pada penelitian ini seperti plotbar, plotpie, ataupun plotline. Matplotlib juga membantu mengatur legend atau keterangan pada data yang disajikan.

- Library Numpy

Numpy digunakan untuk membantu pengelolaan data dalam melempar tipe jenis data array untuk disusun dan diolah karena numpy memiliki kegunaan untuk operasi vektor dan matriks.

Untuk klasifikasian data, digunakan 3 pendekatan berdasarkan jenis data yaitu:

a. Volume trend percakapan

Data yang didapat diolah melalui jumlah data dari populasi data yang dikumpulkan per periode tertentu dengan pengelompokan berdasarkan jenis tren kata kunci yang diekstrak dari questioner DASS-21.

b. Tipe konten isi *tweet*

Data yang diolah berasal dari jumlah konten data yang terbagi menjadi 3 bagian yaitu Reply, Retweet dan Status berdasarkan isi teks dan interaksi antar *username*, setelah dilakukan praproses data. Klasifikasi dihitung berdasarkan isi konten dari tiap kata kunci. Jika diperlukan juga akan dilakukan perhitungan *interaction rate* dimana perhitungannya didapatkan dengan cara nilai retweet dijumlahkan dengan reply dan dibagi jumlah status. ( $\text{Retweet} + \text{Reply} / \text{Status yang dipost}$ ). *Interaction rate* ini kalau di media sosial biasa dikenal dengan *engagement* atau umpan balik. Secara lebih sederhana adalah respon atau perhatian atas suatu status.

c. Pengguna aktif

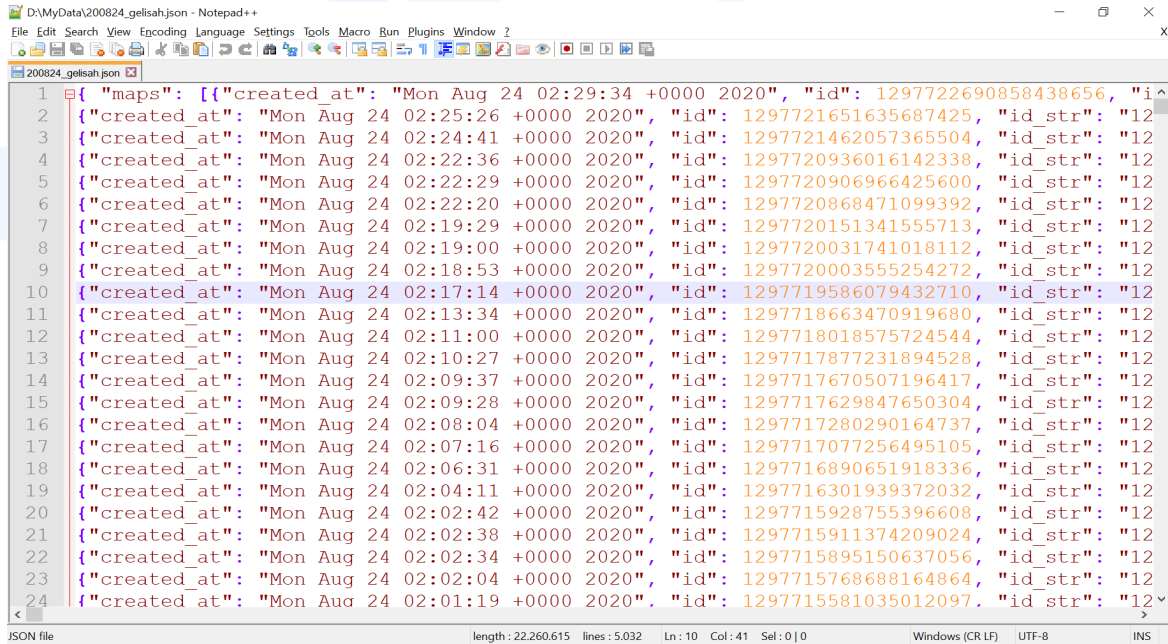
Dalam banyak kasus banyak *buzzer* ataupun *bot* yang turut serta meramaikan suatu topik. Pada kondisi ini penting untuk mengetahui apakah suatu akun itu *real user* atau *bot* agar memudahkan dalam melakukan analisis isi *tweets* yang didapatkan.

## 5.2. Hasil Pengujian Aplikasi Crawler

Aplikasi yang sudah dibangun selanjutnya digunakan untuk menarik data dari Twitter dengan menggunakan beberapa kata kunci yang diekstrak dari DASS-21. Beberapa kata kunci yang digunakan untuk pengetesan aplikasi crawler diantaranya adalah: gelisah, sedih dan

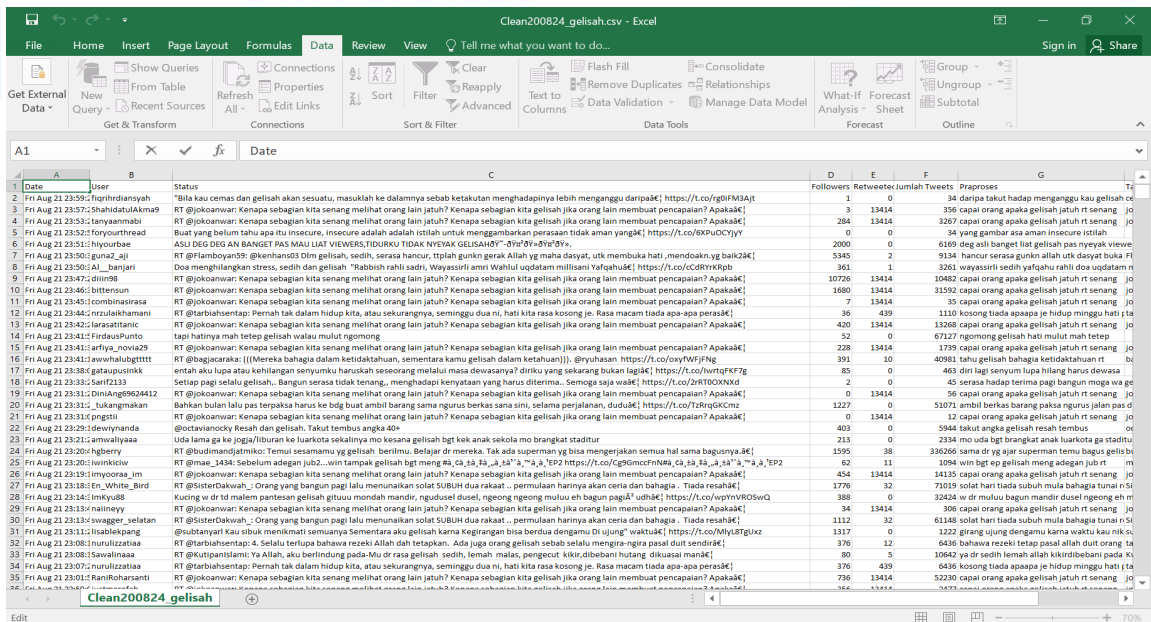
gugup. Informasi lebih lengkap apa saja yang termasuk ke dalam DASS-21 bisa dilihat di Lampiran 1.

Sebagaimana sudah dijelaskan di depan, hasil crawling *tweets* ini dalam bentuk JSON. Berikut ini adalah tampilan JSON yang di dapat dari hasil crawling dengan kata kunci gelisah.



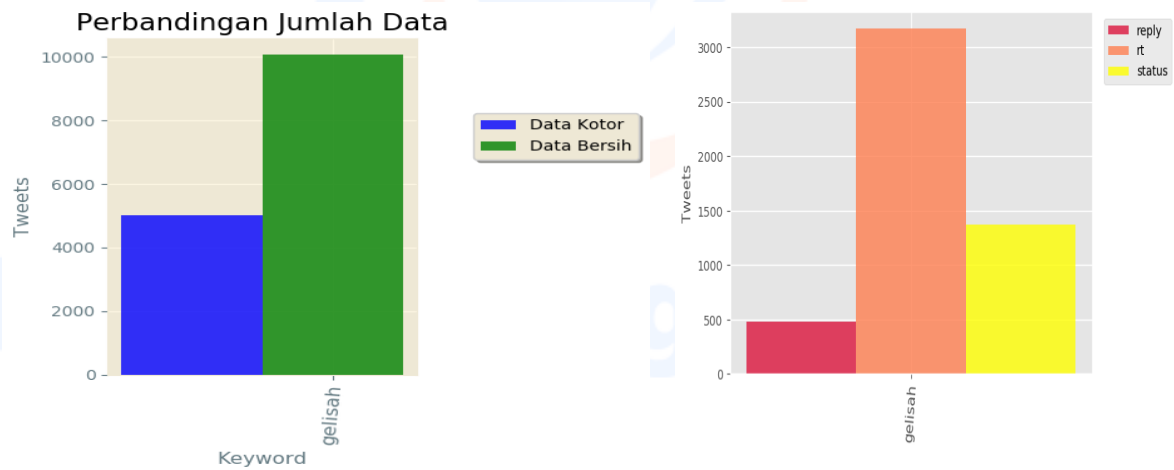
Gambar 5.2. Hasil crawling format JSON dengan kata kunci gelisah

Format JSON ini selanjutnya dilakukan pra-proses yang hasilnya adalah file dengan format .csv. Contoh tampilan format csv yang diperoleh dengan kata kunci gelisah bisa dilihat pada Gambar 5.3

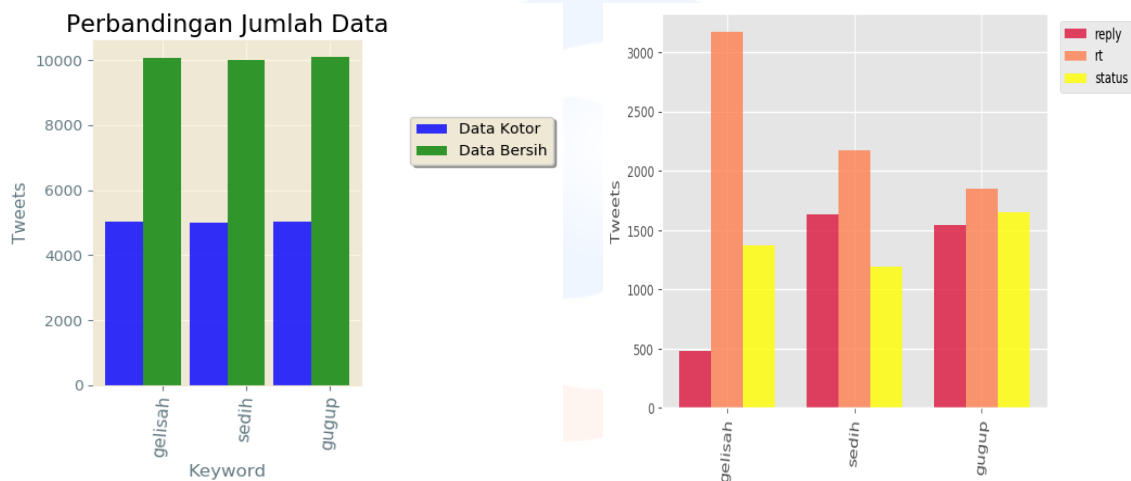


Gambar 5.3. File format csv setelah pra-proses

Selanjutnya hasil pra-proses ini dibuatkan visualisasinya. Hasil visualisasinya dalam bentuk grafik sebagaimana bisa dilihat pada Gambar 5.4.



Gambar 5.4. Hasil visualisasi penarikan data Twitter dengan kata kunci gelisah



Gambar 5.5. Visualisasi perbandingan data Twitter dengan berbagai kata kunci

### 5.3. Pembahasan

Aplikasi crawler yang sudah dibangun sudah bisa dimanfaatkan untuk penarikan data *tweets* dari Twitter. Persoalan mendasarnya adalah penggunaan kata kunci. Penggunaan kata kunci yang kurang tepat akan mengakibatkan data yang ditarik juga kurang sesuai dengan harapan.

Besar bandwidth yang digunakan serta banyaknya data terkait dengan kata kunci yang digunakan sangat mempengaruhi proses penarikan data. Akan lebih baik proses penarikan data dilakukan di luar jam sibuk.

Hingga saat ini Twitter masih memberikan akses gratis untuk penarikan data selama seminggu kemarin. Diluar waktu tersebut dikenakan biaya yang besarnya tergantung kepada

periode yang akan diambil datanya. Untuk kepentingan penelitian ini masih menggunakan akses gratis.

Dari hasil pengujian dengan beberapa kata kunci ini menunjukkan bahwa mesin crawler ini sudah berhasil menarik data dari Twitter secara otomatis. Langkah berikutnya dalam penelitian ini adalah menggunakan mesin *crawler* ini untuk menarik data *tweets* selama beberapa waktu. Tahap berikutnya adalah pembuatan aplikasi *machine learning* untuk klasifikasi dan pemetaan kondisi depresi, kecemasan, dan stress dari hasil postingan sosial media Twitter dari hasil penarikan data ini.

Pada tahap awal ini, aplikasi crawling ini masih berbasis text dan belum memanfaatkan web. Akan lebih baik jika aplikasi ini mulai diintegrasikan ke web sehingga tampilan yang dihasilkan bisa lebih baik dan lebih responsive sehingga bisa diakses dari berbagai media seperti

## BAB VI. KESIMPULAN DAN SARAN

### 6.1. Kesimpulan

Milestone pertama dari penelitian ini adalah pembuatan aplikasi crawler untuk menarik data *tweets* dari media sosial Twitter. Aplikasi crawler yang sudah dibangun berhasil menarik data *tweets* dari Twitter berdasarkan kata kunci tertentu yang diekstrak dari questioner DASS-21 (*Depression Anxiety Stress Scales*).

Hasil dari penarikan data Twitter dengan aplikasi crawler ini bermacam-macam tergantung dari fasenya. Di fase crawling, bisa didapatkan format JSON. Setelah melalui pemrosesan didapatkan data dalam format .csv. Selanjutnya hasilnya di visualisasikan dalam bentuk grafik

### 6.2. Saran

Aplikasi yang sudah dibangun ini masih berbasis teks yang kurang interaktif dan kurang responsive karena di tahap ini masih fokus ke bagaimana bisa menarik data dari Twitter. Dirasa perlu untuk mengintegrasikan aplikasi ini ke web agar bisa lebih interaktif dan responsive sehingga bisa diakses dari berbagai media.

## DAFTAR PUSTAKA

- Bogoroch, R. M. (2005). *Damages for Emotional Distress*.
- Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., & Hanson, C. L. (2016). Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality. *JMIR Mental Health*, 3(2), e21. <https://doi.org/10.2196/mental.4822>
- Cheng, Q., Chang, S. Sen, & Yip, P. S. F. (2012). Opportunities and challenges of online data collection for suicide prevention. *The Lancet*, 379(9830), e53–e54. [https://doi.org/10.1016/S0140-6736\(12\)60856-3](https://doi.org/10.1016/S0140-6736(12)60856-3)
- Cheng, Q., Kwok, C. L., Zhu, T., Guan, L., & Yip, P. S. F. (2015). Suicide communication on social media and its psychological mechanisms: An examination of chinese microblog users. *International Journal of Environmental Research and Public Health*, 12(9), 11506–11527. <https://doi.org/10.3390/ijerph120911506>
- Cheng, Q., Li, T. M., Kwok, C. L., Zhu, T., & Yip, P. S. (2017). Assessing suicide risk and emotional distress in Chinese social media: A text mining and machine learning study. *Journal of Medical Internet Research*, 19(7). <https://doi.org/10.2196/jmir.7276>
- G, M. (2014). 8 shocking suicide attempts posted on the Internet. Retrieved from [http://www.oddee.com/item\\_98907.aspx](http://www.oddee.com/item_98907.aspx)
- Gamon, M., & Counts, S. (2013). Predicting Depression via Social Media. *Comprehensive Child and Adolescent Nursing*, 36(1–2), 168–169. <https://doi.org/10.3109/01460862.2013.798190>
- Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression anxiety stress scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, 44(2), 227–239. <https://doi.org/10.1348/014466505X29657>
- Jarrold, W., Javitz, H. S., Krasnow, R., Peintner, B., Yeh, E., Swan, G. E., & Mehl, M. (2011). Depression and self-focused language in structured interviews with older men. *Psychological Reports*, 109(2), 686–700. <https://doi.org/10.2466/02.09.21.28.PR0.109.5.686-700>
- Lv, M., Li, A., Liu, T., & Zhu, T. (2015). Creating a Chinese suicide dictionary for identifying suicide risk on social media. *PeerJ*, 2015(12). <https://doi.org/10.7717/peerj.1455>
- McCraty, R. . (2006). Emotonal Stress, Positive Emotions, and Psychophysiological Coherence. *Stress in Health and Disease*, 342–365.



- Mirowsky, J., Ross, C. . (2003). *No Title* (2nd Editio). New York: Aldine de Gruyter.
- O'Dea, B., Wan, S., Batterham, P. J., C lear, A. L., Paris, C., & Christensen, H. (2015). Detecting suicidality on twitter. *Internet Interventions*, 2(2), 183–188.  
<https://doi.org/10.1016/j.invent.2015.03.005>
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54(1), 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>
- R.S, L. (1993). From Psychological Stress to the Emotions: A History of Changing Outlook. *Annual Review of Psychology*, 44(1), 21.
- Rickwood, D., Deane, F. P., & Wilson, C. J. (2007). *Getting There : Proto-Psychiatri*. (November).
- Sueki Hajime. (2015). Journal of Affective Disorders. *The Association of Suicide-Related Twitter Use with Suicidal Behaviour: A Cross-Sectional Study of Young Internet Users in Japan*, 170, 155–160. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0165032714005369?via%3Dihub>

# Lampiran 1. DASS-21 (Depression Anxiety Stress Scales)



**DASS 21** NAME \_\_\_\_\_ DATE \_\_\_\_\_

BLACK DOG INSTITUTE

Please read each statement and circle a number 0, 1, 2 or 3 which indicates how much the statement applied to you over the past week. There are no right or wrong answers. Do not spend too much time on any statement.

The rating scale is as follows:

- 0 Did not apply to me at all - NEVER
- 1 Applied to me to some degree, or some of the time - SOMETIMES
- 2 Applied to me to a considerable degree, or a good part of time - OFTEN
- 3 Applied to me very much, or most of the time - ALMOST ALWAYS

					FOR OFFICE USE		
	N	S	O	AA	D	A	S
1	0	1	2	3			
2	0	1	2	3			
3	0	1	2	3			
4	0	1	2	3			
5	0	1	2	3			
6	0	1	2	3			
7	0	1	2	3			
8	0	1	2	3			
9	0	1	2	3			
10	0	1	2	3			
11	0	1	2	3			
12	0	1	2	3			
13	0	1	2	3			
14	0	1	2	3			
15	0	1	2	3			
16	0	1	2	3			
17	0	1	2	3			
18	0	1	2	3			
19	0	1	2	3			
20	0	1	2	3			
21	0	1	2	3			
<b>TOTALS</b>							

## DASS Severity Ratings

The DASS is a **quantitative** measure of distress along the 3 axes of depression, anxiety<sup>1</sup> and stress<sup>2</sup>. It is not a categorical measure of clinical diagnoses.

Emotional syndromes like depression and anxiety are intrinsically dimensional - they vary along a continuum of severity (independent of the specific diagnosis). Hence the selection of a single cut-off score to represent clinical severity is necessarily arbitrary. A scale such as the DASS can lead to a useful assessment of **disturbance**, for example individuals who may fall short of a clinical cut-off for a specific diagnosis can be correctly recognised as experiencing considerable symptoms and as being at high risk of further problems.

However for clinical purposes it can be helpful to have 'labels' to characterise degree of severity relative to the population. Thus the following cut-off scores have been developed for defining mild/moderate/severe/extremely severe scores for each DASS scale.

**Note:** the severity labels are used to describe the full range of scores in the population, so 'mild' for example means that the person is above the population mean but probably still way below the typical severity of someone seeking help (ie it does not mean a mild level of disorder).

The individual DASS scores do not define appropriate interventions. They should be used in conjunction with all clinical information available to you in determining appropriate treatment for any individual.

<sup>1</sup>Symptoms of psychological arousal

<sup>2</sup>The more cognitive, subjective symptoms of anxiety

### DASS 21 SCORE

DEPRESSION SCORE	ANXIETY SCORE	STRESS SCORE
<input type="text"/>	<input type="text"/>	<input type="text"/>

	Depression	Anxiety	Stress
Normal	0 - 4	0 - 3	0 - 7
Mild	5 - 6	4 - 5	8 - 9
Moderate	7 - 10	6 - 7	10 - 12
Severe	11 - 13	8 - 9	13 - 16
Extremely Severe	14 +	10 +	17 +