

Tugas 4 DSA - Albertus Angga Raharja (1606918401, Kelas A)

In [1]:

```
import pandas as pd
import numpy as np
import os
import nltk
# nltk.download('punkt')
# nltk.download('stopwords')
```

In [2]:

```
## Hard coded constants for locating the inputs
PATH_TO_NEW_SENTIMENT_FOLDER = './New Sentiment/Sentiment/'
PATH_TO_DATASET_FILE = PATH_TO_NEW_SENTIMENT_FOLDER + 'data_latih.json'
```

Membaca Data

In [3]:

```
data = pd.read_json(PATH_TO_DATASET_FILE)
data.head()
```

Out[3]:

	akun	id_user	isi	sentimen	tanggal
0	@AhokDjarot	823082326213021696	RT @AhokDjarot: Penasaran sama #BasukiDjarot? ...	netral	Tue Apr 04 14:04:05 +0000 2017
1	@AhokDjarot	3166296157	@AhokDjarot Melayani warga jakarta mulai dari ...	positif	Tue Apr 04 14:00:01 +0000 2017
2	@AhokDjarot	799108167170949121	@Lintank01 @AhokDjarot Golongan sumbu pendek l...	negatif	Tue Apr 04 13:56:37 +0000 2017
3	@AhokDjarot	710566116553166849	RT @RahyaMaya: https://t.co/pvVUIp3YYw Pak Aho...	positif	Tue Apr 04 13:52:44 +0000 2017
4	@AhokDjarot	1335669534	Prof. Edward Omar Sharif H. (Ahli Pidana) : Me...	positif	Tue Apr 04 13:50:01 +0000 2017

In [4]:

```
raw = data['isi'].copy()
raw
```

Out[4]:

```
0    RT @AhokDjarot: Penasaran sama #BasukiDjarot? ...
1    @AhokDjarot Melayani warga jakarta mulai dari ...
2    @Lintank01 @AhokDjarot Golongan sumbu pendek l...
3    RT @RahyaMaya: https://t.co/pvVUIp3YYw Pak Aho...
4    Prof. Edward Omar Sharif H. (Ahli Pidana) : Me...
5    RT @RahyaMaya: https://t.co/pvVUIp3YYw Pak Aho...
6    RT @Jakarta_Kece: Pak Ahok Djarot paling kece ...
7    RT @AhokDjarot: Penasaran sama #BasukiDjarot? ...
8    15. Surah Al Maidah 51 bukan diartikan untuk m...
9    @MudasirRomini @AhokDjarot lah ahok di penjara...
10   RT @RahyaMaya: https://t.co/yxsTUmEql Aku sia...
11   RT @RahyaMaya: https://t.co/pvVUIp3YYw Pak Aho...
12   @Jakarta_Kece @basuki_btp @AhokDjarot @WJLF_ @...
13   RT @Jakarta_Kece: Kita adalah Warga Negara Rep...
14   @Jakarta_Kece @Fakta @AhokDjarot Wkwkwk....ada...
15   @RahyaMaya @Fakta @basuki_btp @AhokDjarot Di d...
16   7. Contoh analoginya ada seorang pencuri di I...
17   @RahyaMaya @Fakta @basuki_btp @AhokDjarot Ini ...
18   RT @Fakta: Kultwit tentang poin-poin Saksi Ahl...
19   RT @Fakta: Kultwit tentang poin-poin Saksi Ahl...
```

```

19 RT @Fakta: Kultwit tentang poin-poin Saksi Ahl...
20 RT @Fakta: Kultwit tentang poin-poin Saksi Ahl...
21 RT @Fakta: Kultwit tentang poin-poin Saksi Ahl...
22 RT @RahyaMaya: https://t.co/owYrQISQG3 Semoga ...
23 RT @RahyaMaya: https://t.co/owYrQISQG3 Semoga ...
24 RT @RahyaMaya: https://t.co/owYrQISQG3 Semoga ...
25 RT @RahyaMaya: https://t.co/owYrQISQG3 Semoga ...
26 RT @Jakarta_Kece: https://t.co/jPaZGq2uhs Subh...
27 Ahok Membangun Masjid @AhokDjarot https://t.co...
28 RT @Jakarta_Kece: Kita adalah Warga Negara Rep...
29 RT @Jakarta_Kece: Kita adalah Warga Negara Rep...
...
620 Strategi Perang Kota @AhokDjarot Makjleb Membe...
621 36. Sudah banyak program @AhokDjarot yg sudah ...
622 Dihantam Banyak Fitnah, Ahok Beri Tahu Relawan...
623 RT @Maya_Rischa: 21. Terbukti, @AhokDjarot buk...
624 soal kualitas, khususnya kualitas pemimpin yg ...
625 RT @SulianQ: Ahok akan Bangun Bioskop Rakyat, ...
626 34. Untuk itu, @AhokDjarot membentuk pasukan k...
627 RT @Maya_Rischa: 29. ....Yang saat ini sdg di...
628 Bioskop Rakyat Dapat Sambutan, Produser Juga I...
629 33. Penataan kota jg menjadi perhatian serius ...
630 Ahok akan Bangun Bioskop Rakyat, Tiketnya? Han...
631 32. ....wisata religi Makam Mbah Priuk, &...
632 RT @Maya_Rischa: 12. Di era @AhokDjarot RSUD d...
633 RT @PaulaAndriani: Djarot: Saya Deg-degan Lho ...
634 Tim @AhokDjarot Sediakan Aplikasi Khusus bagi ...
635 Ahok: Bioskop Akan Ada di Lantai 3 dan 4 Pasar...
636 Ahok dan ide bioskop rakyat di pasar-pasar #Ki...
637 RT @foreveryoudong: 80 Persen Pengguna Faceboo...
638 yg penakut ga cocok jadi gubernur DKI Jakarta ...
639 Padahal rakyat puas dengan kinerja @AhokDjarot...
640 30. Di bidang pariwisata, @AhokDjarot juga tel...
641 . Dia berencana membangun bioskop pribadi mili...
642 RT @Eci_Agustina: Tim @AhokDjarot Gelar Pelati...
643 RT @Eci_Agustina: Tim @AhokDjarot Gelar Pelati...
644 RT @Eci_Agustina: Tim @AhokDjarot Gelar Pelati...
645 RT @Eci_Agustina: Tim @AhokDjarot Gelar Pelati...
646 RT @Eci_Agustina: Tim @AhokDjarot Gelar Pelati...
647 RT @Eci_Agustina: Tim @AhokDjarot Gelar Pelati...
648 RT @Eci_Agustina: Tim @AhokDjarot Gelar Pelati...
649 29. ....Yang saat ini sdg dibangun, adalah bu...
Name: isi, Length: 650, dtype: object

```

Key Norm Dict Untuk Mentransformasi Kata Tidak Baku Menjadi Baku

In [5]:

```

PATH_TO_KEY_NORM_FILE = PATH_TO_NEW_SENTIMENT_FOLDER + 'key_norm.csv'

key_norm = pd.read_csv(PATH_TO_KEY_NORM_FILE).drop(['_id'], axis=1)
key_norm.head()

```

Out [5]:

	singkat	hasil
0	abis	habis
1	accent	tekanan
2	accept	terima
3	accident	kecelakaan
4	achievement	prestasi

In [6]:

```

key_norm_dict = {key_norm['singkat'][i]:key_norm['hasil'][i] for i in range(len(key_norm))}
key_norm_dict

```

Out [6]:

```
{'habis': 'habis'}
```

{
 'axis': 'axis',
 'accent': 'tekanan',
 'accept': 'terima',
 'accident': 'kecelakaan',
 'achievement': 'prestasi',
 'acra': 'acara',
 'acrany': 'acaranya',
 'acrnya': 'acaranya',
 'action': 'aksi',
 'active': 'aktif',
 'activity': 'aktivitas',
 'actually': 'sebenarnya',
 'actualy': 'sebenarnya',
 'ad': 'ada',
 'ade': 'ada',
 'adult': 'dewasa',
 'adventure': 'petualangan',
 'adventurer': 'petualang',
 'advice': 'nasehat',
 'after': 'setelah',
 'afternun': 'sore',
 'again': 'lagi',
 'agency': 'perwakilan',
 'agent': 'agen',
 'agk': 'agak',
 'agktn': 'angkatan',
 'agree': 'setuju',
 'agreement': 'persetujuan',
 'aing': 'saya',
 'aj': 'saja',
 'aja': 'saja',
 'ajah': 'saja',
 'aje': 'saja',
 'ajeh': 'saja',
 'ajk': 'ajak',
 'ak': 'saya',
 'akeh': 'banyak',
 'akhire': 'akhirnya',
 'aktifkn': 'aktifkan',
 'aku': 'saya',
 'alhamdlh': 'alhamdulillah',
 'alhamdulilah': 'alhamdulillah',
 'almost': 'hampir',
 'almt': 'alamat',
 'alone': 'sendiri',
 'alsn': 'alasan',
 'also': 'juga',
 'always': 'selalu',
 'ama': 'sama',
 'amazing': 'mengagumkan',
 'ampe': 'sampai',
 'amu': 'kamu',
 'amuh': 'kamu',
 'ancen': 'memang',
 'ancur': 'hancur',
 'ane': 'saya',
 'anget': 'hangat',
 'aple': 'apple',
 'aplksi': 'aplikasi',
 'aproved': 'diterima',
 'aq': 'saya',
 'aqoe': 'saya',
 'aqu': 'saya',
 'aqw': 'saya',
 'arep': 'mau',
 'arived': 'tiba',
 'article': 'artikel',
 'artine': 'artinya',
 'assignment': 'tugas',
 'assistance': 'bantuan',
 'assistant': 'asisten',
 'asyek': 'asyik',
 'ats': 'atas',
 'atu': 'satu-satu',
 'atw': 'atau',
 'automatic': 'otomatis',
 'aya': 'ada',
 'b': 'bb'

u : u ,
'b\\kesinambungn": 'berkesinambungan',
'babe": 'ayah',
'balasanya': 'balasannya',
'bales': 'balas',
'banga': 'bangga',
'batre': 'baterai',
'bbrapa': 'beberapa',
'bbrp': 'beberapa',
'bbrpa': 'beberapa',
'bdg': 'bandung',
'beach': 'pantai',
'before': 'sebelum',
'begete': 'banget',
'bego': 'bodoh',
'belike': 'belikan',
'bengi': 'malam',
'benjol': 'memar',
'beraktifitas': 'beraktivitas',
'berjln': 'berjalan',
'berkmmampuan': 'berkemampuan',
'bersaman': 'bersamaan',
'bet': 'banget',
'beter': 'better',
'between': 'antara',
'Bgm': 'bagaimana',
'bgmn': 'bagaimana',
'bgmna': 'bagaimana',
'bgn': 'bangun',
'bgt': 'banget',
'bgtu': 'begitu',
'bhs': 'bahasa',
'bimbel': 'bimbingan belajar',
'birthday': 'ulang tahun',
'bis': 'habis',
'bjln': 'berjalan',
'bkln': 'bakalan',
'bkn': 'bukan',
'blackberry': 'blackberry',
'blg': 'bilang',
'blk': 'balik',
'blm': 'belum',
'bln': 'bulan',
'blng': 'bilang',
'blnja': 'belanja',
'blom': 'belum',
'blon': 'belum',
'bls': 'balas',
'blsanya': 'balasannya',
'blum': 'belum',
'bm': 'bbm',
'bnar': 'benar',
'bner': 'benar',
'bnr': 'benar',
'bnrn': 'beneran',
'bntr': 'sebentar',
'bntu': 'bantu',
'bnuh': 'bunuh',
'bnyk': 'banyak',
'bobonya': 'tidurnya',
'bodo': 'bodoh',
'bokap': 'ayah',
'bole': 'boleh',
'boong': 'bohong',
'boring': 'bosan',
'bpak': 'bapak',
'bpk': 'bapak',
'brangcut': 'berangkat',
'brangkt': 'berangkat',
'brenang': 'berenang',
'brg': 'barang',
'brp': 'berapa',
'brpa': 'berapa',
'brsih': 'bersih',
'bru': 'baru',
'bs': 'bisa',
'bsa': 'bisa',
'bsa': 'bisa'

'besar': 'besar',
'bsh': 'basah',
'bsk': 'besok',
'bsok': 'besok',
'bsr': 'besar',
'bth': 'butuh',
'btuh': 'butuh',
'btul': 'betul',
'btw': 'ngomong ngomong',
'buanged': 'banget',
'buangedh': 'banget',
'buanget': 'banget',
'buktiin': 'buktikan',
'buktin': 'buktikan',
'buln': 'bulan',
'bungks': 'bungkus',
'bw': 'bawa',
'bwa': 'bawa',
'bwah': 'bawah',
'bwat': 'untuk',
'bwh': 'bawah',
'bwt': 'untuk',
'byar': 'bayar',
'byk': 'banyak',
'byr': 'bayar',
'cabe': 'cabai',
'cadas': 'cerdas',
'cah': 'bocah',
'call': 'telepon',
'can': 'dapat',
'cape': 'capek',
'car\r\n': 'mobil',
'career': 'karir',
'caution': 'perhatian',
'cb': 'coba',
'Ceman': 'teman',
'cemungudh': 'semangat',
'Cemungut': 'semangat',
'cepat': 'cepat',
'cew': 'cewek',
'cewe': 'cewek',
'check': 'cek',
'ciapa': 'siapa',
'cibuk': 'sibuk',
'cinto': 'cinta',
'cius': 'serius',
'ciusan': 'seriusan',
'ciyus': 'serius',
'clananya': 'celananya',
'cmpakan': 'campakkan',
'cmpakin': 'campakkan',
'cnd': 'canda',
'cnda': 'canda',
'cnggih': 'canggih',
'cngih': 'canggih',
'cnt': 'cinta',
'cnta': 'cinta',
'cntik': 'cantik',
'cntk': 'cantik',
'cntre': 'tengah',
'co': 'cowok',
'coach': 'pelatih',
'coaster': 'alas',
'cocokin': 'cocokkan',
'cocoknyo': 'cocoknya',
'coconut': 'kelapa',
'cod': 'bayar saat pengiriman',
'coding': 'mengkode',
'coklat': 'cokelat',
'colecting': 'mengoleksi',
'colection': 'koleksi',
'colongin': 'curikan',
'colonial': 'kolonial',
'color': 'warna',
'colorful': 'berwarna-warni',
'colour': 'warna',
'comeback': 'datang kembali',
'comedy': 'komedi',

comedy': 'komedi',
'comendable': 'patut dihargai',
'coment': 'komen',
'comercial': 'komersil',
'comes': 'datang',
'confirm': 'konfirmasi',
'Comfort': 'nyaman',
'comfortable': 'menyenangkan',
'comic': 'komik',
'coming': 'datang',
'comingson': 'segera datang',
'comingsoon': 'segera datang',
'commando': 'prajurit',
'commendable': 'patut dihargai',
'comment': 'komen',
'commentary': 'uraian',
'commercial': 'komersil',
'community': 'komunita',
'communityny': 'komunitanya',
'company': 'perusahaan',
'competition': 'persaingan',
'complain': 'mengeluh',
'complete': 'lengkap',
'completes': 'lengkap',
'compliment': 'pujian',
'compose': 'menyusun',
'computing': 'menghitung',
'community': 'komunita',
'communityny': 'komunitanya',
'concept': 'konsep',
'concert': 'konser',
'condition': 'kondisi',
'conditioning': 'mengkondisikan',
'conect': 'menghubungkan',
'conection': 'koneksi',
'confident': 'yakin',
'configuration': 'susunan',
'confirm': 'konfirmasi',
'confius': 'bingung',
'confuse': 'bingung',
'confused': 'bingung',
'congralsts': '(ucapan) selamat',
'Congrat': '(ucapan) selamat',
'congrate': '(ucapan) selamat',
'congrats': '(ucapan) selamat',
'consept': 'konsep',
'consider': 'mempertimbangkan',
'consult': 'berunding',
'consultants': 'konsultan',
'consumer': 'konsumen',
'contac': 'kontak',
'contact': 'kontak',
'content': 'isi',
'contest': 'kontes',
'contoe': 'contohnya',
'contra': 'kontra',
'contract': 'kontrak',
'control': 'kontrol',
'convention': 'konvensi',
'Convoy': 'konvoi',
'cook': 'memasak',
'copas': 'copy paste',
'copot': 'lepas',
'corner': 'sudut',
'corners': 'sudut',
'corola': 'corolla',
'correction': 'koreksi',
'corridor': 'koridor',
'cosmetics': 'kosmetik',
'cost': 'ongkos',
'costs': 'ongkos',
'costum': 'kostum',
'could': 'dapat',
'council': 'dewan',
'countdown': 'perhitungan detik-detik terakhir',
'counter': 'alat perhitungan',
'country': 'negara',

countryside': 'daerah pedalaman',
 couple': 'pasangan',
 cousin': 'sepupu',
 cow': 'cowok',
 cowo': 'cowok',
 cpt': 'cepat',
 cr': 'cara',
 crazy': 'gila',
 cri': 'cari',
 crt': 'cerita',
 crtX': 'ceritanya',
 cucok': 'cocok',
 cukup': 'cukup',
 cuman': 'cuma',
 cumn': 'hanya',
 curcol': 'curhat',
 cwek': 'cewek',
 cwo': 'cowok',
 cwok': 'cowok',
 cyus': 'serius',
 dah': 'sudah',
 dalan': 'jalan',
 dangerous': 'bahaya',
 dasyat': 'dahsyat',
 date': 'tanggal',
 dateng': 'datang',
 datanglah': 'datanglah',
 day': 'hari',
 dbls': 'dibalas',
 dbnt': 'dibantu',
 ddlm': 'didalam',
 dduk': 'duduk',
 defence': 'pertahanan',
 defend': 'membela',
 deg2an': 'deg-degan',
 degdeg': 'deg-degan',
 degdeg-an': 'deg-degan',
 degdegan': 'deg-degan',
 degdegser': 'deg-degan',
 deket': 'dekat',
 desek2an': 'desak-desakan',
 desekan': 'desakan',
 dewe': 'sendiri',
 dftar': 'daftar',
 dg': 'dengan',
 dgn': 'dengan',
 dgr': 'dengar',
 dgrin': 'dengarkan',
 dgrn': 'dengarkan',
 dh': 'sudah',
 dibangain': 'dibanggakan',
 dibls': 'dibalas',
 digangu': 'diganggu',
 diizinin': 'diberi izin',
 dikasi': 'dikasih',
 dikasinya': 'dikasihnya',
 dikit': 'sedikit',
 dikls': 'di kelas',
 diliat': 'dilihat',
 diluk': 'sebentar',
 dimana2': 'dimana-mana',
 dimkn': 'dimakan',
 dimksd': 'dimaksud',
 dimn2': 'dimana-mana',
 dimna': 'dimana',
 dinstal': 'diinstall',
 dipake': 'dipakai',
 dipangilnya': 'dipanggilnya',
 dipke': 'dipakai',
 disawat': 'dilempar',
 disblah': 'disebelah',
 discon': 'diskon',
 discont': 'diskon',
 discount': 'diskon',
 disebrang': 'diseberang',
 disebuah': 'disebuah',
 disekolah': 'di sekolah',

'disekolahan': 'di sekolah',
'disekolahin': 'disekolahkan',
'diselenggarakan': 'diselenggarakan',
'diselimutin': 'diselimuti',
'disenayan': 'di senayan',
'disenen': 'di senen',
'disentul': 'di sentul',
'diservice': 'diservis',
'dishub': 'Dinas Perhubungan',
'dishubkominformasi': 'Dinas Perhubungan Komunikasi dan Informatika',
'disi': 'diisi',
'disia2in': 'disia-siakan',
'disimpen': 'disimpan',
'disimpng': 'disimpang',
'disinih': 'disini',
'disklh': 'di sekolah',
'diskonnye': 'diskonnya',
'diskony': 'diskonnya',
'diskonya': 'diskonnya',
'Diskonye': 'diskonnya',
'disni': 'disini',
'disono': 'disana',
'disonoh': 'disana',
'distribution': 'distribusi',
'district': 'daerah',
'distroy': 'hancur',
'disturb': 'ganggu',
'disurga': 'di surga',
'disurh': 'disuruh',
'disuru': 'disuruh',
'ditabra': 'ditabrak',
'ditaktir': 'ditraktir',
'ditaman': 'di taman',
'ditambahin': 'ditambahkan',
'ditanah': 'di tanah',
'ditangepin': 'ditanggapin',
'ditanggepin': 'ditanggapin',
'ditangkep': 'ditangkap',
'ditangkepinnya': 'ditangkapinnya',
'ditangkepinya': 'ditangkapinnya',
'ditanggung': 'ditanggung',
'ditaro': 'ditaruh',
'ditarok': 'ditaruh',
'ditelen': 'ditelan',
'ditemeni': 'ditemani',
'ditemenin': 'ditemani',
'diteraktir': 'ditraktir',
'ditingal': 'ditinggal',
'ditingali': 'ditinggali',
'ditingikan': 'ditinggikan',
'ditkarkan': 'ditukarkan',
'ditlp': 'ditelepon',
'ditmpt': 'di tempat',
'ditngah': 'ditengah',
'ditngguin': 'ditungguin',
'ditngu': 'ditunggu',
'ditnjkn': 'ditunjukkan',
'ditoko': 'di toko',
'ditrakter': 'ditraktir',
'ditrotoar': 'di trotoar',
'ditugu': 'di tugu',
'dituker': 'ditukar',
'ditukoke': 'dibelikan',
'ditukokno': 'dibelikan',
'ditumbasne': 'dibelikan',
'ditumpaki': 'dinaiki',
'diTunas': 'di tunas',
'ditunggu2': 'ditunggu-tunggu',
'ditungu': 'ditunggu',
'ditungu2': 'ditunggu-tunggu',
'ditunguin': 'ditunggu',
'ditunjukan': 'ditunjukkan',
'ditunjukin': 'ditunjukkan',
'diturunkn': 'diturunkan',
'diwarnet': 'di warnet',
'diweekend': 'di akhir pekan',
'diwolak': 'dibalik',

'djagain': 'dijaga',
'djemput': 'dijemput',
'dJkt': 'di Jakarta',
'djln': 'di jalan',
'dkasih': 'dikasih',
'dkat': 'dekat',
'dkatkan': 'dekatkan',
'dket': 'dekat',
'dketan': 'dekat',
'dkit': 'sedikit',
'dkk': 'dan kawan-kawan',
'dklinikk': 'di klinik',
'dkluarin': 'dikeluarkan',
'dkmps': 'di kampus',
'dkomplain': 'dikomplain',
'dksh': 'dikasih',
'dkt': 'dekat',
'dkung': 'dukung',
'dl': 'dulu',
'dlakukan': 'dilakukan',
'dlakukn': 'dilakukan',
'dlam': 'dalam',
'dlanjutin': 'dilanjutkan',
'dliat': 'dilihat',
'dliburkan': 'diliburkan',
'dll': 'dan lain-lain',
'dlm': 'dalam',
'dlnjutin': 'dilanjutkan',
'dlu': 'dulu',
'dluan': 'dulu',
'dmainkan': 'dimainkan',
'dmakan': 'dimakan',
'dmam': 'demam',
'dmamX': 'demamnya',
'dman': 'dimana',
'dmana': 'dimana',
'dminta': 'diminta',
'dmn': 'dimana',
'dmna': 'dimana',
'dmna2': 'dimana-mana',
'dmnanya': 'dimananya',
'dmnnya': 'dimananya',
'dmno': 'dimana',
'dmnt': 'diminta',
'dmnya': 'dimananya',
'dmobil': 'di mobil',
'dmonas': 'di monas',
'dmtr': 'di motor',
'dn': 'dan',
'dng': 'dengan',
'dngan': 'dengan',
'dnger': 'dengar',
'dnger2': 'denga-dengar',
'dngerin': 'dengarkan',
'dnggo': 'dipakai',
'dngin': 'dingin',
'dnginan': 'dinginan',
'dngn': 'dengan',
'doain': 'doakan',
'doane': 'doanya',
'doang': 'saja',
'doank': 'saja',
'dolali': 'pada lupa',
'dolan': 'bermain',
'donload': 'unduh',
'donlod': 'unduh',
'donlot': 'unduh',
'dont': 'jangan',
'dpat': 'dapat',
'dpatkan': 'dapatkan',
'dpet': 'dapat',
'dpn': 'depan',
'dpstikn': 'dipastikan',
'dpt': 'dapat',
'dr': 'dari',
'dri': 'dari',
'dripada': 'daripada',

'dripd': 'daripada',
'drmh': 'di rumah',
'drmn': 'darimana',
'drpd': 'daripada',
'drttd': 'daritadi',
'dsana': 'disana',
'dsar': 'dasar',
'dsb': 'dan sebagainya',
'dsini': 'disini',
'dsitu': 'disitu',
'dskolah': 'di sekolah',
'dsktr': 'di sekitar',
'dsmb': 'desember',
'dsna': 'disana',
'dsni': 'disini',
'dsno': 'disana',
'dsruh': 'disuruh',
'dst': 'dan seterusnya',
'dtang': 'datang',
'dteng': 'datang',
'dtg': 'datang',
'dtgin': 'datangin',
'dturun': 'diturunkan',
'duitna': 'duitnya',
'during': 'selama',
'duwe': 'punya',
'duwit': 'duit',
'dy': 'dia',
'elo': 'kamu',
'elu': 'kamu',
'emg': 'emang',
'emng': 'emang',
'emoh': 'tidak mau',
'endiri': 'sendiri',
'endonesia': 'Indonesia',
'enelaan': 'beneran',
'engak': 'tidak',
'engga': 'tidak',
'enggak': 'tidak',
'engk': 'tidak',
'enk': 'enak',
'enough': 'cukup',
'entar': 'nanti',
'enteng': 'ringan',
'enyak': 'enak',
'etc': 'dan lain-lain',
'event': 'acara',
'every': 'setiap',
'everybody': 'semua orang',
'everything': 'segala sesuatu',
'everywhere': 'dimanapun',
'ex': 'contoh',
'excellent': 'bagus',
'exit': 'keluar',
'face': 'muka',
'fact': 'fakta',
'failed': 'gagal',
'father': 'ayah',
'fee': 'biaya',
'ferari': 'ferrari',
'filem': 'film',
'finally': 'akhirnya',
'focus': 'fokus',
'folback': 'follow back',
'folbacknya': 'follow back nya',
'folbc': 'follow back',
'folbcak': 'follow back',
'folbek': 'follow back',
'folbeknya': 'follow back nya',
'follback': 'follow back',
'follbacknya': 'follow back nya',
'follbc': 'follow back',
'follo': 'follow',
'followbak': 'follow back',
'folo': 'follow',
'follow': 'follow',
'followback': 'follow back',

'followbak': 'follow back',
'followbcak': 'follow back',
'football': 'sepak bola',
'for': 'untuk',
'forged': 'lupa',
'forget': 'lupa',
'forgot': 'lupa',
'foto2': 'foto-foto',
'friday': 'jumat',
'friend': 'teman',
'frofil': 'profil',
'from': 'dari',
'function': 'fungsi',
'funny': 'lucu',
'funtion': 'fungsi',
'funy': 'lucu',
'g': 'tidak',
'ga': 'tidak',
'gaada': 'tidak ada',
'gaakan': 'tidak akan',
'gaasik': 'tidak asik',
'gabakal': 'tak akan',
'gabareng': 'tidak bareng',
'gabebas': 'tidak bebas',
'gabeda': 'tidak beda',
'gaberani': 'tidak berani',
'gabilang': 'tidak bilang',
'gabisa': 'tidak bisa',
'gabole': 'tidak boleh',
'gaboleh': 'tidak boleh',
'gabosen': 'tidak bosen',
'gabs': 'tidak bisa',
'gad': 'tidak ada',
'gada': 'tidak ada',
'gadaiin': 'gadaikan',
'Gadapet': 'tidak dapat',
'gadein': 'gadaikan',
'gadeket': 'tidak dekat',
'gadiajak': 'tidak diajak',
'gadibales': 'tidak dibalas',
'gae': 'membuat',
'gaean': 'pekerjaan',
'gaenak': 'tidak enak',
'gaene': 'biasanya',
'gag': 'tidak',
'gahoel': 'gaul',
'gahol': 'gaul',
'gaikut': 'tidak ikut',
'gaikutaaan': 'tidak ikutan',
'gaikutan': 'tidak ikutan',
'gaiso': 'tidak bisa',
'gajadi': 'tidak jadi',
'gajalan': 'tidak jalan',
'gajanji': 'tidak janji',
'gajauh': 'tidak jauh',
'Gajd': 'tidak jadi',
'Gaje': 'tidak jelas',
'gajelas': 'tidak jelas',
'gajenguk': 'tidak menjenguk',
'gak': 'tidak',
'gakada': 'tidak ada',
'Gakan': 'tidak akan',
'gakbilang': 'tidak bilang',
'gakbisa': 'tidak bisa',
'gakdeh': 'tidak deh',
'gakenak': 'tidak enak',
'gakjadi': 'tidak jadi',
'gakmau': 'tidak mau',
'gakpapa': 'tidak apa-apa',
'gakpenting': 'tidak penting',
'gakpernah': 'tidak pernah',
'gaksabar': 'tidak sabar',
'gaksemua': 'tidak semua',
'gaksih': 'tidak sih',
'gaktau': 'tidak tau',
'gaktuh': 'tidak tuh',
'gamain': 'tidak main',

'gamales': 'tidak malas',
'gamao': 'tidak mau',
'gamasuk': 'tidak masuk',
'gamau': 'tidak mau',
'gamsk': 'tidak masuk',
'gamuat': 'tidak muat',
'gamungkin': 'tidak mungkin',
'gan': 'juragan',
'gangajak': 'tidak mengajak',
'Gangerti': 'tidak mengerti',
'ganguin': 'gangguin',
'gantiin': 'gantikan',
'gantinen': 'ganti saja',
'gaosah': 'tidak usah',
'gapada': 'tidak pada',
'gapaham': 'tidak paham',
'gapake': 'tidak pakai',
'gapapa': 'tidak apa-apa',
'gapercaya': 'tidak percaya',
'gaperlu': 'tidak perlu',
'gapernah': 'tidak pernah',
'gapo': 'tidak apa-apa',
'gaptek': 'gagap teknologi',
'gapulang': 'tidak pulang',
'gara2': 'gara-gara',
'gasabar': 'tidak sabar',
'gasalah': 'tidak salah',
'gasanggup': 'tidak sanggup',
'gasecerah': 'tidak secerah',
'gasedingin': 'tidak sedingin',
'gasekeras': 'tidak sekeras',
'gaseru': 'tidak seru',
'gasia2': 'tidak sia-sia',
'gasih': 'tidak sih',
'gasik': 'datang lebih awal',
'gasuka': 'tidak suka',
'gatau': 'tidak tau',
'gatel': 'gatal',
'gathel': 'gatal',
'gatuh': 'tidak tuh',
'Gausa': 'tidak usah',
'gausah': 'tidak usah',
'gaush': 'tidak usah',
'gawake': 'bawakan',
'gawakne': 'bawakan',
'gayalu': 'gayamu',
'gbawa': 'tidak bawa',
'gbk': 'Gelora Bung Karno',
'gblg': 'tidak bilang',
'gblk': 'tidak balik',
'gbs': 'tidak bisa',
'gbsa': 'tidak bisa',
'gda': 'tidak ada',
'gebetan': 'pasangan',
'gebetn': 'pasangan',
'geblek': 'bodoh',
'gede': 'besar',
'gedenya': 'besarnya',
'gedhe2an': 'besar-besaran',
'gelem': 'bersedia',
'gelo': 'gila',
'gendeng': 'bodoh',
'gendheng': 'bodoh',
'gendud': 'gemuk',
'geto': 'begitu',
'getoh': 'begitu',
'gg': 'tidak',
'gilak': 'gila',
'gile': 'gila',
'gimana': 'bagaimana',
'gimane': 'bagaimana',
'gimna': 'bagaimana',
'gini': 'begini',
'gitu': 'begitu',
'gituan': 'begituan',
'gituh': 'begitu',
'gjd': 'tidak jadi',

'gjdi': 'tidak jadi',
'gjls': 'tidak jelas',
'gjnji': 'tidak janji',
'gk': 'tidak',
'gmbr': 'gambar',
'gmn': 'bagaimana',
'gmna': 'bagaimana',
'gmne': 'bagaimana',
'gnti': 'ganti',
'goceng': 'lima ribu',
'godan': 'godaan',
'good luck': 'semoga berhasil',
'goodmorning': 'selamat pagi',
'goro2': 'gara-gara',
'gpl': 'tidak pakai lama',
'gpp': 'tidak-apa-apa',
'gpplah': 'tidak apa-apalah',
'gr2': 'gara-gara',
'gra': 'gara-gara',
'grade': 'tingkat',
'gt': 'begitu',
'gtu': 'begitu',
'gtw': 'tidak tau',
'gua': 'saya',
'gue': 'saya',
'gueh': 'saya',
'guweh': 'saya',
'gw': 'saya',
'gwe': 'saya',
'gx': 'tidak',
'gy': 'sedang',
'hany': 'hanya',
'hape': 'hp',
'hapi': 'senang',
'happy': 'senang',
'happybday': 'selamat ulang tahun',
'hareem': 'perempuan',
'harg': 'harga',
'harganya': 'harganya',
'hargonyo': 'harganya',
'hasek': 'asyik',
'hawong': 'lhawong',
'hbd': 'selamat ulang tahun',
'hbs': 'habis',
'hdiah': 'hadiah',
'hdiahnya': 'hadiahnya',
'hdirkan': 'hadirkan',
'hdp': 'hidup',
'hdpn': 'hadapan',
'hdup': 'hidup',
'heng': 'error',
'hidp': 'hidup',
'hidpmu': 'hidupmu',
'hinga': 'hingga',
'hmpas': 'hempas',
'hmpir': 'hampir',
'hntu': 'hantu',
'hny': 'hanya',
'hnya': 'hanya',
'Hoam': 'ngantuk',
'honest': 'jujur',
'hr': 'hari',
'hrapan': 'harapan',
'hrg': 'harga',
'hrga': 'harga',
'hrgnya': 'harganya',
'hri': 'hari',
'hrs': 'harus',
'hrsnya': 'harusnya',
'hrus': 'harus',
'hrz': 'harus',
'hsil': 'hasil',
'hsl': 'hasil',
'htelnya': 'hotelnya',
'htiku': 'hatiku',
'hub': 'hubungi',
'hungry': 'lapar',

'ibdah': 'ibadah',
'idung': 'hidung',
'ijin': 'izin',
'ijo': 'hijau',
'iki': 'ini',
'ikt': 'ikut',
'iktan': 'ikutan',
'iktn': 'ikutan',
'ikud': 'ikut',
'ilang': 'hilang',
'img': 'gambar',
'inbox': 'kotak masuk',
'inceran': 'incaran',
'include': 'termasuk',
'indh': 'indah',
'indo': 'indonesia',
'info': 'informasi',
'infonya': 'informasinya',
'information': 'informasi',
'informkan': 'informasikan',
'infox': 'informasinya',
'inget': 'ingat',
'inginya': 'inginnya',
'ingt': 'ingat',
'ingtn': 'ingatan',
'inovation': 'inovasi',
'inpo': 'informasi',
'inpoh': 'informasi',
'insaallah': 'insyaAllah',
'insyaa': 'insyaAllah',
'insyaaloh': 'insyaAllah',
'insyah': 'insyaAllah',
'insyalah': 'insyaAllah',
'intrnet': 'internet',
'invit': 'mengajak',
'invite': 'mengajak',
'invte': 'mengajak',
'io': 'iya',
'ireng': 'hitam',
'iseh': 'masih',
'isih': 'masih',
'isin': 'malu',
'isine': 'isinya',
'iso': 'bisa',
'isteri': 'istri',
'istrht': 'istirahat',
'isuk': 'pagi',
'itung': 'hitung',
'iy': 'iya',
'iyadah': 'iya dah',
'iyadeh': 'iya dah',
'iyadong': 'iya dong',
'iyadund': 'iya dong',
'iyah': 'iya',
'iyanih': 'iya nih',
'iyasih': 'iya sih',
'iyaw': 'iya',
'iye': 'iya',
'iyedah': 'iya dah',
'iyelah': 'iyalah',
'iyeu': 'iya',
'iyh': 'iya',
'iyo': 'iya',
'iyosih': 'iya sih',
'iyuh': 'iya',
'iyuwih': 'iya',
'j': 'saja',
'ja': 'saja',
'jabar': 'Jawa Barat',
'jadina': 'jadinya',
'jadinge': 'jadinya',
'jadul': 'jaman dulu',
'jadwalnye': 'jadwalnya',
'jahate': 'jahatnya',
'jakpus': 'Jakarta Pusat',
'jakrta': 'Jakarta',
'jaksel': 'Jakarta Selatan',

```
'jaktim': 'Jakarta Timur',
'jalan2': 'jalan-jalan',
'jalankn': 'jalankan',
'jalanya': 'jalannya',
'jamanna': 'jamannya',
'jamanny': 'jamannya',
'jamany': 'jamannya',
'jamanya': 'jamannya',
'jamanye': 'jamannya',
'jangn': 'jangan',
'japan': 'jepang',
'jare': 'kata',
'jaringanx': 'jaringannya',
'jaringn': 'jaringan',
'jateng': 'Jawa Tengah',
'jatoh': 'jatuh',
'jaz': 'jazz',
'jd': 'jadi',
'jdi': 'jadi',
'jdian': 'jadian',
'jenenge': 'namanya',
'jenengi': 'beri nama',
'jep': 'jeep',
'jg': 'juga',
'jga': 'juga',
'jgja': 'jogja',
'jgn': 'jangan',
'jgn2': 'jangan-jangan',
'jgnkan': 'jangankan',
'jht': 'jahat',
'jjur': 'jujur',
'jk': 'jika',
'jkrt': 'Jakarta',
'jkrta': 'Jakarta',
'jkt': 'Jakarta',
'jl': 'jalan',
'jlan2': 'jalan-jalan',
'jlas': 'jelas',
'jln': 'jalan',
'jln2': 'jalan-jalan',
'jls': 'jelas',
'jmbatan': 'jembatan',
'jmbtan': 'jembatan',
'jmin': 'jamin',
'jmlh': 'jumlah',
'jmn': 'jaman',
'jmpa': 'jumpa',
'jmpe': 'jumpa',
'jmpt': 'jemput',
'jmptnya': 'jemputnya',
'jo': 'jangan',
'jobseeker': 'pencari kerja',
'jri': 'jari',
'jringan': 'jaringan',
'jringn': 'jaringan',
'jrngn': 'jaringan',
'jtuh': 'jatuh',
'jua': 'juga',
'jug': 'juga',
'jugaj': 'juga',
'jugak': 'juga',
'jugo': 'juga',
'jwb': 'jawab',
'jwban': 'jawaban',
'jwbanny': 'jawabannya',
'jwbannya': 'jawabannya',
'jwbany': 'jawabannya',
'kabare': 'kabarnya',
'kabeh': 'semua',
'kacian': 'kasihan',
'kadieu': 'kesini',
'kaditu': 'kesitu',
'kaga': 'tidak',
'kagak': 'tidak',
'kagk': 'tidak',
'kakean': 'kebanyakan',
...}
```

Membaca Custom Stop Word List `stopword_list_TALA.txt`

In [7]:

```
PATH_TO_STOP_WORD_LIST_FILE = PATH_TO_NEW_SENTIMENT_FOLDER + 'stopword_list_TALA.txt'

with open(PATH_TO_STOP_WORD_LIST_FILE, 'r') as stop_word_list_file:
    stop_words = [word.rstrip('\n').lower() for word in stop_word_list_file.readlines()]

stop_words
```

Out[7]:

```
['at_user',
 'atuser',
 'url',
 'rt',
 'ada',
 'adalah',
 'adanya',
 'adapun',
 'agak',
 'agaknya',
 'agar',
 'akan',
 'akankah',
 'akhir',
 'akhiri',
 'akhirnya',
 'aku',
 'akulah',
 'amat',
 'amatlah',
 'anda',
 'andalah',
 'antar',
 'antara',
 'antaranya',
 'apa',
 'apaan',
 'apabila',
 'apakah',
 'apalagi',
 'apatah',
 'artinya',
 'asal',
 'asalkan',
 'atas',
 'atau',
 'ataukah',
 'ataupun',
 'awal',
 'awalnya',
 'bagai',
 'bagaikan',
 'bagaimana',
 'bagaimanakah',
 'bagaimanapun',
 'bagi',
 'bagian',
 'bahkan',
 'bahwa',
 'bahwasanya',
 'baik',
 'bakal',
 'bakalan',
 'balik',
 'banyak',
 'bapak',
 'baru',
 'bawah',
 'beberapa',
 'begini',
```


'beginian',
'beginikah',
'beginilah',
'begitu',
'begitukah',
'begitulah',
'begitupun',
'bekerja',
'belakang',
'belakangan',
'belum',
'belumlah',
'benar',
'benarkah',
'benarlah',
'berada',
'berakhir',
'berakhirlah',
'berakhirnya',
'berapa',
'berapakah',
'berapalah',
'berapapun',
'berarti',
'berawal',
'berbagai',
'berdatangan',
'beri',
'berikan',
'berikut',
'berikutnya',
'berjumlah',
'berkali-kali',
'berkata',
'berkehendak',
'berkeinginan',
'berkenaan',
'berlainan',
'berlalu',
'berlangsung',
'berlebihan',
'bermacam',
'bermacam-macam',
'bermaksud',
'bermula',
'bersama',
'bersama-sama',
'bersiap',
'bersiap-siap',
'bertanya',
'bertanya-tanya',
'berturut',
'berturut-turut',
'bertutur',
'berujar',
'berupa',
'besar',
'betul',
'betulkah',
'biasa',
'biasanya',
'bila',
'bilakah',
'bisa',
'bisakah',
'boleh',
'bolehkah',
'bolehlah',
'buat',
'bukan',
'bukankah',
'bukanlah',
'bukannya',
'bulan',
'bung',
'cara',
'caranva',

'cukup',
'cukupkah',
'cukuplah',
'cuma',
'dahulu',
'dalam',
'dan',
'dapat',
'dari',
'daripada',
'datang',
'dekat',
'demi',
'demikian',
'demikianlah',
'dengan',
'depan',
'di',
'dia',
'diakhiri',
'diakhirinya',
'dialah',
'diantara',
'diantaranya',
'diberi',
'diberikan',
'diberikannya',
'dibuat',
'dibuatnya',
'didapat',
'didatangkan',
'digunakan',
'diibaratkan',
'diibaratkannya',
'diingat',
'diingatkan',
'diinginkan',
'dijawab',
'dijelaskan',
'dijelaskannya',
'dikarenakan',
'dikatakan',
'dikatakannya',
'dikerjakan',
'diketahui',
'diketahuinya',
'dikira',
'dilakukan',
'dilalui',
'dilihat',
'dimaksud',
'dimaksudkan',
'dimaksudkannya',
'dimaksudnya',
'diminta',
'dimintai',
'dimisalkan',
'dimulai',
'dimulailah',
'dimulainya',
'dimungkinkan',
'dini',
'dipastikan',
'diperbuat',
'diperbuatnya',
'dipergunakan',
'diperkirakan',
'diperlihatkan',
'diperlukan',
'diperlukannya',
'dipersoalkan',
'dipertanyakan',
'dipunyai',
'diri',
'dirinya',
'disampaikan',
'disebut'.

'disebutkan',
'disebutkannya',
'disini',
'disinilah',
'ditambahkan',
'ditandaskan',
'ditanya',
'ditanyai',
'ditanyakan',
'ditegaskan',
'ditujukan',
'ditunjuk',
'ditunjuki',
'ditunjukkan',
'ditunjukkannya',
'ditunjuknya',
'dituturkan',
'dituturkannya',
'diucapkan',
'diucapkannya',
'diungkapkan',
'dong',
'dua',
'dulu',
'empat',
'enggak',
'enggaknya',
'entah',
'entahlah',
'guna',
'gunakan',
'hal',
'hampir',
'hanya',
'hanyalah',
'hari',
'harus',
'haruslah',
'harusnya',
'hendak',
'hendaklah',
'hendaknya',
'hingga',
'ia',
'ialah',
'ibarat',
'ibaratkan',
'ibaratnya',
'ibu',
'ikut',
'ingat',
'ingat-ingat',
'ingin',
'inginkah',
'inginkan',
'ini',
'inikah',
'inilah',
'itu',
'itukah',
'itulah',
'jadi',
'jadilah',
'jadinya',
'jangan',
'janganakan',
'janganlah',
'jauh',
'jawab',
'jawaban',
'jawabnya',
'jelas',
'jelaskan',
'jelaslah',
'jelasnya',
'jika',
'ikalan'.

'juga',
'jumlah',
'jumlahnya',
'justru',
'kala',
'kalau',
'kalaulah',
'kalaupun',
'kalian',
'kami',
'kamilah',
'kamu',
'kamulah',
'kan',
'kapan',
'kapankah',
'kapanpun',
'karena',
'karenanya',
'kasus',
'kata',
'katakan',
'katakanlah',
'katanya',
'ke',
'keadaan',
'kebetulan',
'kecil',
'kedua',
'keduanya',
'keinginan',
'kelamaan',
'kelihatan',
'kelihatannya',
'kelima',
'keluar',
'kembali',
'kemudian',
'kemungkinan',
'kemungkinannya',
'kenapa',
'kepada',
'kepadanya',
'kesampaian',
'keseluruhan',
'keseluruhannya',
'keterlaluhan',
'ketika',
'khususnya',
'kini',
'kinilah',
'kira',
'kira-kira',
'kiranya',
'kita',
'kitalah',
'kok',
'kurang',
'lagi',
'lagian',
'lah',
'lain',
'lainnya',
'lalu',
'lama',
'lamanya',
'lanjut',
'lanjutnya',
'lebih',
'lewat',
'lima',
'luar',
'macam',
'maka',
'makanya',
'makin',
'malah'

malah,
'malahan',
'mampu',
'mampukah',
'mana',
'manakala',
'manalagi',
'masa',
'masalah',
'masalahnya',
'masih',
'masihkah',
'masing',
'masing-masing',
'mau',
'maupun',
'melainkan',
'melakukan',
'melalui',
'melihat',
'melihatnya',
'memang',
'memastikan',
'memberi',
'memberikan',
'membuat',
'memerlukan',
'memihak',
'meminta',
'memintakan',
'memisalkan',
'memperbuat',
'mempergunakan',
'memperkirakan',
'memperlihatkan',
'mempersiapkan',
'mempersoalkan',
'mempertanyakan',
'mempunyai',
'memulai',
'memungkinkan',
'menaiki',
'menambahkan',
'menandaskan',
'menanti',
'menanti-nanti',
'menantikan',
'menanya',
'menanyai',
'menanyakan',
'mendapat',
'mendapatkan',
'mendatang',
'mendatangi',
'mendatangkan',
'menegaskan',
'mengakhiri',
'mengapa',
'mengatakan',
'mengatakannya',
'mengenai',
'mengerjakan',
'mengetahui',
'menggunakan',
'menghendaki',
'mengibaratkan',
'mengibaratkannya',
'mengingat',
'mengingatkan',
'menginginkan',
'mengira',
'mengucapkan',
'mengucapkannya',
'mengungkapkan',
'menjadi',
'menjawab',
'menjelaskan',
'menyebut'

menuju',
'menunjuk',
'menunjuki',
'menunjukkan',
'menunjuknya',
'menurut',
'menuturkan',
'menyampaikan',
'menyangkut',
'menyatakan',
'menyebutkan',
'menyeluruh',
'menyiapkan',
'merasa',
'mereka',
'merekalah',
'merupakan',
'meski',
'meskipun',
'meyakini',
'meyakinkan',
'minta',
'mirip',
'misal',
'misalkan',
'misalnya',
'mula',
'mulai',
'mulailah',
'mulanya',
'mungkin',
'mungkinkah',
'nah',
'naik',
'namun',
'nanti',
'nantinya',
'nyaris',
'nyatanya',
'oleh',
'olehnya',
'pada',
'padahal',
'padanya',
'pak',
'paling',
'panjang',
'pantas',
'para',
'pasti',
'pastilah',
'penting',
'pentingnya',
'per',
'percuma',
'perlu',
'perlukah',
'perlunya',
'pernah',
'persoalan',
'pertama',
'pertama-tama',
'pertanyaan',
'pertanyakan',
'pihak',
'pihaknya',
'pukul',
'pula',
'pun',
'punya',
'rasa',
'rasanya',
'rata',
'rupanya',
'saat',
'saatnya',
'saja',
'sebaliknya'

'saja',
'saling',
'sama',
'sama-sama',
'sambil',
'sampai',
'sampai-sampai',
'sampaikan',
'sana',
'sangat',
'sangatlah',
'satu',
'saya',
'sayalah',
'se',
'sebab',
'sebabnya',
'sebagai',
'sebagaimana',
'sebagainya',
'sebagian',
'sebaik',
'sebaik-baiknya',
'sebaiknya',
'sebaliknya',
'sebanyak',
'sebegini',
'sebegitu',
'sebelum',
'sebelumnya',
'sebenarnya',
'seberapa',
'sebesar',
'sebetulnya',
'sebisanya',
'sebuah',
'sebut',
'sebutlah',
'sebutnya',
'secara',
'secukupnya',
'sedang',
'sedangkan',
'sedemikian',
'sedikit',
'sedikitnya',
'seenaknya',
'segala',
'segalanya',
'segera',
'seharusnya',
'sehingga',
'seingat',
'sejak',
'sejauh',
'sejenak',
'sejumlah',
'sekadar',
'sekadarnya',
'sekali',
'sekali-kali',
'sekalian',
'sekaligus',
'sekalipun',
'sekarang',
'sekarang',
'sekecil',
'seketika',
'sekiranya',
'sekitar',
'sekitarnya',
'sekurang-kurangnya',
'sekurangnya',
'sela',
'selain',
'selaku',
'selalu',
'selama',

'selama',
'selama-lamanya',
'selamanya',
'selanjutnya',
'seluruh',
'seluruhnya',
'semacam',
'semakin',
'semampu',
'semampunya',
'semasa',
'semasih',
'semata',
'semata-mata',
'semaunya',
'sementara',
'semisal',
'semisalnya',
'sempat',
'semua',
'semuanya',
'semula',
'sendiri',
'sendirian',
'sendirinya',
'seolah',
'seolah-olah',
'seorang',
'sepanjang',
'sepantasnya',
'sepantasnyalah',
'seperlunya',
'seperti',
'sepertinya',
'sepihak',
'sering',
'seringnya',
'serta',
'serupa',
'sesaat',
'sesama',
'sesampai',
'sesegera',
'sesekali',
'seseorang',
'sesuatu',
'sesuatunya',
'sesudah',
'sesudahnya',
'setelah',
'setempat',
'setengah',
'seterusnya',
'setiap',
'setiba',
'setibanya',
'setidak-tidaknya',
'setidaknya',
'setinggi',
'seusai',
'sewaktu',
'siap',
'siapa',
'siapakah',
'siapapun',
'sini',
'sinilah',
'soal',
'soalnya',
'suatu',
'sudah',
'sudahkah',
'sudahlah',
'supaya',
'tadi',
'tadinya',
'tahu',
'...'

'tahun',
'tak',
'tambah',
'tambahnya',
'tampak',
'tampaknya',
'tandas',
'tandasnya',
'tanpa',
'tanya',
'tanyakan',
'tanyanya',
'tapi',
'tegas',
'tegasnya',
'telah',
'tempat',
'tengah',
'tentang',
'tentu',
'tentulah',
'tentunya',
'tepat',
'terakhir',
'terasa',
'terbanyak',
'terdahulu',
'terdapat',
'terdiri',
'terhadap',
'terhadapnya',
'teringat',
'teringat-ingat',
'terjadi',
'terjadilah',
'terjadinya',
'terkira',
'terlalu',
'terlebih',
'terlihat',
'termasuk',
'ternyata',
'tersampaikan',
'tersebut',
'tersebutlah',
'tertentu',
'tertuju',
'terus',
'terutama',
'tetap',
'tetapi',
'tiap',
'tiba',
'tiba-tiba',
'tidak',
'tidakkah',
'tidaklah',
'tiga',
'tinggi',
'toh',
'tunjuk',
'turut',
'tutur',
'tuturnya',
'ucap',
'ucapnya',
'ujar',
'ujarnya',
'umum',
'umumnya',
'ungkap',
'ungkapnya',
'untuk',
'usah',
'usai',
'waduh',
'wah',
..

```
'wahai',
'waktu',
'waktunya',
'walau',
'walaupun',
'wong',
'yaitu',
'yakin',
'yakni',
'yang']
```

Basic Preprocessing

In [113]:

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemover, ArrayDictionary
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from nltk.tokenize import word_tokenize

import re

# Create StopWord Removal using our own Stop words
stop_word_remover = StopWordRemover(ArrayDictionary(stop_words))

# Create stemmer
stemmer = StemmerFactory().create_stemmer()

def formalize(word):
    if word in key_norm_dict:
        return key_norm_dict[word]
    return word

def preprocess_tweet(tweet, debug = False):
    if debug: print(tweet)

    # Lower casing
    clean_tweet = tweet.lower() #lowercase

    # URL Removal and Username removal
    clean_tweet = re.sub(r"(?:\@|https?\:\/\/)\S+", " ", clean_tweet)
    if debug: print('After URL and username removal: ', clean_tweet)

    # Punctuation removal
    clean_tweet = re.sub(r'^\w\s|', ' ', clean_tweet)
    if debug: print('After punctuation removal: ', clean_tweet)

    # Extra space removal
    clean_tweet = re.sub('\s+', ' ', clean_tweet)
    if debug: print('After extra space removal: ', clean_tweet)

    # Trimming
    clean_tweet = clean_tweet.strip()
    if debug: print('After trimming: ', clean_tweet)

    # Stop words removal
    clean_tweet = stop_word_remover.remove(clean_tweet)
    if debug: print('After stop words removal: ', clean_tweet)

    # Transforming informal words to formal words
    clean_tweet = " ".join([formalize(word) for word in word_tokenize(clean_tweet)])
    if debug: print('After transofrming informal words: ', clean_tweet)

    # Repeating characters removal
    copy_clean_tweet = clean_tweet
    rpt_regex = re.compile(r"(.)\1{1,}", re.IGNORECASE) #regex for repeating characters
    clean_tweet = re.sub(rpt_regex, '', clean_tweet)
    if copy_clean_tweet != clean_tweet:
        print('After repeating characters removal: ', clean_tweet)

    # Stemming
    clean_tweet = stemmer.stem(clean_tweet)
    if debug: print('After stemming: ', clean_tweet)

    return clean_tweet
```

In [114]:

```
cleaned = pd.Series([preprocess_tweet(tweet) for tweet in data['isi']])
data['isi'] = cleaned
data
```

Out[114]:

	akun	id_user	isi	sentimen	tanggal
0	@AhokDjarot	823082326213021696	penasaran basukidjarot pertany nkepo kepoinpel...	netral	Tue Apr 04 14:04:05 +0000 2017
1	@AhokDjarot	3166296157	layan warga jakarta lahir program fr ahok	positif	Tue Apr 04 14:00:01 +0000 2017
2	@AhokDjarot	799108167170949121	golong sumbu pendek fpi piara keluarga cendana...	negatif	Tue Apr 04 13:56:37 +0000 2017
3	@AhokDjarot	710566116553166849	ahok didzolimi ahok bebas fr ahok	positif	Tue Apr 04 13:52:44 +0000 2017
4	@AhokDjarot	1335669534	prof edward omar sharif h ahli pidana pndpt ah...	positif	Tue Apr 04 13:50:01 +0000 2017
5	@AhokDjarot	2357765268	ahok didzolimi ahok bebas fr ahok	positif	Tue Apr 04 13:43:59 +0000 2017
6	@AhokDjarot	3970978754	ahok djarot gaul badai tata kota jakarta fr ahok	positif	Tue Apr 04 13:43:02 +0000 2017
7	@AhokDjarot	60637169	penasaran basukidjarot pertany nkepo kepoinpel...	netral	Tue Apr 04 13:41:17 +0000 2017
8	@AhokDjarot	1335669534	15 surah al maidah 51 arti umat pilih pimpin m...	positif	Tue Apr 04 13:41:01 +0000 2017
9	@AhokDjarot	1222354812	ahok penjara bersih2 penjara biar ahok	negatif	Tue Apr 04 13:39:44 +0000 2017
10	@AhokDjarot	545368722	tun program un ul ahok djarot fr ahok	positif	Tue Apr 04 13:36:59 +0000 2017
11	@AhokDjarot	545368722	ahok didzolimi ahok bebas fr ahok	positif	Tue Apr 04 13:36:49 +0000 2017
12	@AhokDjarot	794537177472303104	ik t	positif	Tue Apr 04 13:35:44 +0000 2017
13	@AhokDjarot	794537177472303104	warga negara republik indonesia sosok ahok fr ...	positif	Tue Apr 04 13:35:33 +0000 2017
14	@AhokDjarot	230151878	wkwkwk udang dbalik batu	negatif	Tue Apr 04 13:34:27 +0000 2017
15	@AhokDjarot	230151878	dzolimi bgmana jela mendolimi umat islam sampe...	negatif	Tue Apr 04 13:33:24 +0000 2017
16	@AhokDjarot	1335669534	7 contoh analogi curi indonesia dasar hukum po...	positif	Tue Apr 04 13:33:02 +0000 2017
17	@AhokDjarot	230151878	nista agama y	negatif	Tue Apr 04 13:32:04 +0000 2017
18	@AhokDjarot	806368329124474880	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:30:42 +0000 2017
19	@AhokDjarot	2958782773	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:30:42 +0000 2017
20	@AhokDjarot	3328346382	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:30:42 +0000 2017
21	@AhokDjarot	807132193881346048	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:30:42 +0000 2017
22	@AhokDjarot	3328346382	moga ahok djarot un ul putar fr ahok	positif	Tue Apr 04 13:30:14 +0000 2017
23	@AhokDjarot	1335669534	moga ahok djarot un ul putar fr ahok	positif	Tue Apr 04 13:30:14 +0000 2017
24	@AhokDjarot	807132193881346048	moga ahok djarot un ul putar fr ahok	positif	Tue Apr 04 13:30:14 +0000 2017
25	@AhokDjarot	2958782773	moga ahok djarot un ul putar fr ahok	positif	Tue Apr 04 13:30:14 +0000 2017
26	@AhokDjarot	795293967185100800	subhana ah djarot fr ahok	positif	Tue Apr 04 13:29:23 +0000 2017
27	@AhokDjarot	2207235865	ahok bangun masjid fr ahok	positif	Tue Apr 04 13:28:01 +0000 2017
28	@AhokDjarot	2958782773	warga negara republik indonesia sosok ahok fr ...	positif	Tue Apr 04 13:26:59 +0000 2017
29	@AhokDjarot	807132193881346048	warga negara republik indonesia sosok ahok fr ...	positif	Tue Apr 04 13:26:59 +0000 2017
...
620	@AhokDjarot	2207309154	strategi perang kota makjleb bendung operasi g...	positif	Mon Apr 03 12:36:01 +0000 2017
621	@AhokDjarot	808168296860127233	36 program bukti kerja nyata bukti manf t war	netral	Mon Apr 03 12:36:01 +0000 2017
622	@AhokDjarot	597598439	hantam fitnah ahok rawan tangkis kita2badja	netral	Mon Apr 03 12:36:00 +0000 2017
623	@AhokDjarot	2901856978	21 bukti bangun fisik jakarta mental warga ban...	positif	Mon Apr 03 12:35:52 +0000 2017
624	@AhokDjarot	2207200386	kualitas kualitas pimpin pilih gubernur dki	positif	Mon Apr 03 12:35:01 +0000 2017
625	@AhokDjarot	839351851824762880	ahok bangun bioskop rakyat tiket rp10 ribu kit...	netral	Mon Apr 03 12:34:12 +0000 2017
626	@AhokDjarot	808168296860127233	34 bentuk pasu khusus biru oranye kuning hijau...	netral	Mon Apr 03 12:34:01 +0000 2017

627	@AhokDjarot	170955597	29 bangun bukti nyata macet kita2badja	isi sentimen	Mon Apr 03 12:33:55 +0000 2017
628	@AhokDjarot	597598439	bioskop rakyat sambut produser film eks kita2b...	netral	Mon Apr 03 12:33:01 +0000 2017
629	@AhokDjarot	808168296860127233	tata kota perhati serius era bersih indah amp ...	netral	Mon Apr 03 12:33:01 +0000 2017
630	@AhokDjarot	597598439	ahok bangun bioskop rakyat tiket rp10 ribu kit...	netral	Mon Apr 03 12:32:01 +0000 2017
631	@AhokDjarot	808168296860127233	32 wisata religi makam mbah priuk amp keliling...	netral	Mon Apr 03 12:32:01 +0000 2017
632	@AhokDjarot	822579060547158017	12 era rsud kualitas tara rumah sakit swasta n...	netral	Mon Apr 03 12:31:59 +0000 2017
633	@AhokDjarot	812494141040967680	djarot deg degan lho ahok bilang pecat kita2badja	negatif	Mon Apr 03 12:31:19 +0000 2017
634	@AhokDjarot	2207309154	tim sedia aplikasi khusus saksi tps kita2badja	netral	Mon Apr 03 12:31:01 +0000 2017
635	@AhokDjarot	812984602545373184	ahok bioskop lantai 3 4 pasar pasar jakarta ki...	netral	Mon Apr 03 12:31:01 +0000 2017
636	@AhokDjarot	597598439	ahok ide bioskop rakyat pasar pasar kita2badja	netral	Mon Apr 03 12:31:01 +0000 2017
637	@AhokDjarot	812494141040967680	80 persen faceb k twi er suka ahok kita2badja	positif	Mon Apr 03 12:30:42 +0000 2017
638	@AhokDjarot	805714772171685888	takut cocok gubernur dki jakarta ya om m f om	negatif	Mon Apr 03 12:30:03 +0000 2017
639	@AhokDjarot	166100153	rakyat puas kerja ubah jakarta signifikan kita...	positif	Mon Apr 03 12:30:03 +0000 2017
640	@AhokDjarot	808168296860127233	30 bidang pariwisata bukti buka wisata kita2badja	netral	Mon Apr 03 12:30:01 +0000 2017
641	@AhokDjarot	597649776	rencana bangun bioskop pribadi milik perintah ...	netral	Mon Apr 03 12:30:01 +0000 2017
642	@AhokDjarot	810139401242214402	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
643	@AhokDjarot	777456541335035904	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
644	@AhokDjarot	809783453428264960	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
645	@AhokDjarot	821663582882500608	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
646	@AhokDjarot	810136877256863746	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
647	@AhokDjarot	821695975064412160	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
648	@AhokDjarot	1270831718	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
649	@AhokDjarot	808168296860127233	29 bangun bukti nyata macet kita2badja	positif	Mon Apr 03 12:29:01 +0000 2017

650 rows × 5 columns

Komparasi setelah dan sebelum preprocessing

In [115]:

```
raw.head()
```

Out[115]:

```
0    RT @AhokDjarot: Penasaran sama #BasukiDjarot? ...
1    @AhokDjarot Melayani warga jakarta mulai dari ...
2    @Lintank01 @AhokDjarot Golongan sumbu pendek l...
3    RT @RahyaMaya: https://t.co/pvVUIp3YYw Pak Aho...
4    Prof. Edward Omar Sharif H. (Ahli Pidana) : Me...
Name: isi, dtype: object
```

In [116]:

```
cleaned.head()
```

Out[116]:

```
0    penasaran basukidjarot pertany nkepo kepoinpel...
1        layan warga jakarta lahir program fr ahok
2    golong sumbu pendek fpi piara keluarga cendana...
3        ahok didzolimi ahok bebas fr ahok
4    prof edward omar sharif h ahli pidana pndpt ah...
dtype: object
```

In [117]:

```
for tweet in data['isi']:
    for word in word_tokenize(tweet):
        if word in stop_words:
            print(tweet, word)
```

2 - Word Cloud untuk setiap kelas Sentimen

In [118]:

```
from wordcloud import WordCloud
%matplotlib inline
import matplotlib.pyplot as plt

def plot_word_cloud(data):
    wordcloud = WordCloud().generate(" ".join(data['isi']))
    plt.figure(figsize = (16, 8))
    plt.imshow(wordcloud, interpolation = 'bilinear')
    plt.axis("off")
    plt.show()
```

Nilai Sentimen pada Data

In [119]:

```
set(data['sentimen'])
```

Out[119]:

```
{'negatif', 'netral', 'positif'}
```

Word Cloud untuk Kelas Sentimen = Positif

In [120]:

```
positif_data = data[data['sentimen'] == 'positif']
positif_data
```

Out[120]:

	akun	id_user	isi	sentimen	tanggal
1	@AhokDjarot	3166296157	layan warga jakarta lahir program fr ahok	positif	Tue Apr 04 14:00:01 +0000 2017
3	@AhokDjarot	710566116553166849	ahok didzolimi ahok bebas fr ahok	positif	Tue Apr 04 13:52:44 +0000 2017
4	@AhokDjarot	1335669534	prof edward omar sharif h ahli pidana pndpt ah...	positif	Tue Apr 04 13:50:01 +0000 2017
5	@AhokDjarot	2357765268	ahok didzolimi ahok bebas fr ahok	positif	Tue Apr 04 13:43:59 +0000 2017
6	@AhokDjarot	3970978754	ahok djarot gaul badai tata kota jakarta fr ahok	positif	Tue Apr 04 13:43:02 +0000 2017
8	@AhokDjarot	1335669534	15 surah al maidah 51 arti umat pilih pimpin m...	positif	Tue Apr 04 13:41:01 +0000 2017
10	@AhokDjarot	545368722	tun program un ul ahok djarot fr ahok	positif	Tue Apr 04 13:36:59 +0000 2017
11	@AhokDjarot	545368722	ahok didzolimi ahok bebas fr ahok	positif	Tue Apr 04 13:36:49 +0000 2017

id	username	id_user	isi	sentimen	tanggal
12	@AhokDjarot	794537177472303104		positif	Tue Apr 04 13:35:44 +0000 2017
13	@AhokDjarot	794537177472303104	warga negara republik indonesia sosok ahok fr ...	positif	Tue Apr 04 13:35:33 +0000 2017
16	@AhokDjarot	1335669534	7 contoh analogi curi indonesia dasar hukum po...	positif	Tue Apr 04 13:33:02 +0000 2017
18	@AhokDjarot	806368329124474880	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:30:42 +0000 2017
19	@AhokDjarot	2958782773	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:30:42 +0000 2017
20	@AhokDjarot	3328346382	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:30:42 +0000 2017
21	@AhokDjarot	807132193881346048	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:30:42 +0000 2017
22	@AhokDjarot	3328346382	moga ahok djarot un ul putar fr ahok	positif	Tue Apr 04 13:30:14 +0000 2017
23	@AhokDjarot	1335669534	moga ahok djarot un ul putar fr ahok	positif	Tue Apr 04 13:30:14 +0000 2017
24	@AhokDjarot	807132193881346048	moga ahok djarot un ul putar fr ahok	positif	Tue Apr 04 13:30:14 +0000 2017
25	@AhokDjarot	2958782773	moga ahok djarot un ul putar fr ahok	positif	Tue Apr 04 13:30:14 +0000 2017
26	@AhokDjarot	795293967185100800	subhana ah djarot fr ahok	positif	Tue Apr 04 13:29:23 +0000 2017
27	@AhokDjarot	2207235865	ahok bangun masjid fr ahok	positif	Tue Apr 04 13:28:01 +0000 2017
28	@AhokDjarot	2958782773	warga negara republik indonesia sosok ahok fr ...	positif	Tue Apr 04 13:26:59 +0000 2017
29	@AhokDjarot	807132193881346048	warga negara republik indonesia sosok ahok fr ...	positif	Tue Apr 04 13:26:59 +0000 2017
30	@AhokDjarot	1335669534	warga negara republik indonesia sosok ahok fr ...	positif	Tue Apr 04 13:26:59 +0000 2017
31	@AhokDjarot	806368329124474880	warga negara republik indonesia sosok ahok fr ...	positif	Tue Apr 04 13:26:59 +0000 2017
32	@AhokDjarot	806368329124474880	ahok djarot gaul badai tata kota jakarta fr ahok	positif	Tue Apr 04 13:26:29 +0000 2017
33	@AhokDjarot	2958782773	ahok djarot gaul badai tata kota jakarta fr ahok	positif	Tue Apr 04 13:26:29 +0000 2017
34	@AhokDjarot	807132193881346048	ahok djarot gaul badai tata kota jakarta fr ahok	positif	Tue Apr 04 13:26:29 +0000 2017
35	@AhokDjarot	1335669534	ahok djarot gaul badai tata kota jakarta fr ahok	positif	Tue Apr 04 13:26:29 +0000 2017
36	@AhokDjarot	1335669534	kultwit poin poin saksi ahli sidang ahok yuk s...	positif	Tue Apr 04 13:26:01 +0000 2017
...
510	@AhokDjarot	1349434237	jakarta creative hubung kumpul kreatif cipta p...	positif	Mon Apr 03 13:47:55 +0000 2017
511	@AhokDjarot	1349434237	transjakarta angkutan2 janji bantu remaja angk...	positif	Mon Apr 03 13:47:50 +0000 2017
517	@AhokDjarot	1349434237	ikhlas layan warga jakarta basuki t purnama ja...	positif	Mon Apr 03 13:47:34 +0000 2017
522	@AhokDjarot	1349434237	70 puas masyarakat djarot berterimakasih	positif	Mon Apr 03 13:47:06 +0000 2017
523	@AhokDjarot	1349434237	dukung basuki djarot patungan terdft amp paka...	positif	Mon Apr 03 13:47:02 +0000 2017
529	@AhokDjarot	845074626044616704	mantap maju	positif	Mon Apr 03 13:46:38 +0000 2017
531	@AhokDjarot	794946245986951168	awesome	positif	Mon Apr 03 13:42:15 +0000 2017
555	@AhokDjarot	3283345260	kerja team sukses	positif	Mon Apr 03 13:19:38 +0000 2017
562	@AhokDjarot	808251067922161664	djarot blusukan pasar pagi nyonya kita2badja	positif	Mon Apr 03 13:14:04 +0000 2017
563	@AhokDjarot	810495542983847937	kita2badja semangat badja rawan gentar masuk k...	positif	Mon Apr 03 13:14:01 +0000 2017
568	@AhokDjarot	805714772171685888	mantap jiwa dukung penjur tanah air untuk kit...	positif	Mon Apr 03 13:10:01 +0000 2017
585	@AhokDjarot	837277243277639682	program kerja basuki dj	positif	Mon Apr 03 13:02:49 +0000 2017
598	@AhokDjarot	2207200386	ahok djarot sempurna untung bonus malam	positif	Mon Apr 03 12:56:01 +0000 2017
599	@AhokDjarot	812494141040967680	dukung warga madura jakarta kualitas integrita...	positif	Mon Apr 03 12:55:11 +0000 2017
600	@AhokDjarot	599143758	nusron wahid ajak warga madura jakarta pilih a...	positif	Mon Apr 03 12:55:02 +0000 2017
601	@AhokDjarot	2207200386	ahok djarot bonus duga kita2badja	positif	Mon Apr 03 12:54:01 +0000 2017
602	@AhokDjarot	2540610164	keren lurah estate manger ide brilian ahok dja...	positif	Mon Apr 03 12:50:19 +0000 2017
603	@AhokDjarot	777485683971596288	kita2badja ajak warga nyanyi cbr aksi djarot p...	positif	Mon Apr 03 12:49:01 +0000 2017

tanqqal

0	@AhokDjarot	823082326213621696	penasaran basukidjarot pertany nkepo kepoinpel...	netral	Tue Apr 04 14:04:05 +0000 2017
7	@AhokDjarot	60637169	penasaran basukidjarot pertany nkepo kepoinpel...	netral	Tue Apr 04 13:41:17 +0000 2017
74	@AhokDjarot	2883237643	ahokdjarotdirosi warga dki harap pilkada dki pu	netral	Tue Apr 04 12:46:31 +0000 2017
76	@AhokDjarot	799108167170949121	ahok tkw saudi kerja bersih bersih masjid 1 ko...	netral	Tue Apr 04 12:38:30 +0000 2017
90	@AhokDjarot	1363774536	ya kembar wakil gubernur	netral	Tue Apr 04 11:02:02 +0000 2017
97	@AhokDjarot	1363604977	lombamiripdjarot	netral	Tue Apr 04 10:54:26 +0000 2017
99	@AhokDjarot	815843569193783296	seru nih lomba lombamiripdjarot	netral	Tue Apr 04 10:52:45 +0000 2017
102	@AhokDjarot	827412451213746176	mbah sampean kali kawan saudara ikutin lombami...	netral	Tue Apr 04 10:31:43 +0000 2017
103	@AhokDjarot	827461839827005440	ayo slankers kali lombamiripdjarot	netral	Tue Apr 04 10:29:57 +0000 2017
105	@AhokDjarot	841558591500767234	om2 nyok lombamiripdjarot	netral	Tue Apr 04 10:15:46 +0000 2017
106	@AhokDjarot	841560757187702785	buru kirim foto om kawan saudara lombamiripdjarot	netral	Tue Apr 04 10:14:23 +0000 2017
107	@AhokDjarot	1266741102	kumis kumis whatsa kesini 085 7392681 hadiah s...	netral	Tue Apr 04 10:14:18 +0000 2017
108	@AhokDjarot	841565632504053761	kumis ayo lombamiripdjarot	netral	Tue Apr 04 10:11:17 +0000 2017
109	@AhokDjarot	1266799196	kumis djarot whatsa kesini 085 7392681 hadiah ...	netral	Tue Apr 04 10:10:13 +0000 2017
110	@AhokDjarot	841567802800529408	hadiah ayo lombamiripdjarot	netral	Tue Apr 04 10:09:57 +0000 2017
111	@AhokDjarot	819461489312419840	ayo lombamiripdjarot	netral	Tue Apr 04 10:08:35 +0000 2017
123	@AhokDjarot	2919157490	om swastyastu selamat raya nyepi hening moga t...	netral	Tue Apr 04 08:22:36 +0000 2017
124	@AhokDjarot	2919157490	1 jakartapunyasemua fasilitas tunjang bangsa h...	netral	Tue Apr 04 08:22:23 +0000 2017
134	@AhokDjarot	168643229	duka pulang ayahanda senin 3 april 2017 moga d...	netral	Tue Apr 04 07:48:21 +0000 2017
144	@AhokDjarot	167898289	duka pulang ayahanda senin 3 april 2017 moga d...	netral	Tue Apr 04 06:35:28 +0000 2017
149	@AhokDjarot	2929102448	jakarta rusun festival rebut hadiah utama unju...	netral	Tue Apr 04 05:58:42 +0000 2017
151	@AhokDjarot	951491120	aniestakutdebat	netral	Tue Apr 04 05:35:33 +0000 2017
155	@AhokDjarot	2943350389	duka pulang ayahanda senin 3 april 2017 moga d...	netral	Tue Apr 04 04:57:50 +0000 2017
156	@AhokDjarot	2943350389	pagi suara adzan sentuh kalbu lonceng gereja d...	netral	Tue Apr 04 04:56:37 +0000 2017
163	@AhokDjarot	49589450	jakarta rusun festival rebut hadiah utama unju...	netral	Tue Apr 04 04:02:44 +0000 2017
184	@AhokDjarot	808251067922161664	rptra kalijodo bangun ahok djarot maskot rptra...	netral	Tue Apr 04 03:08:02 +0000 2017
207	@AhokDjarot	4688622475	1 program pimpin jakarta kita2badja	netral	Tue Apr 04 02:24:15 +0000 2017
224	@AhokDjarot	834228353104171009		netral	Tue Apr 04 02:12:22 +0000 2017
245	@AhokDjarot	815273617474752512		netral	Tue Apr 04 01:01:47 +0000 2017
248	@AhokDjarot	398798880	program jad n warung jakmart akses langsung pu...	netral	Tue Apr 04 00:40:31 +0000 2017
...
593	@AhokDjarot	779151018386857984	bangun fisik rumah sakit rptra rumah ibadah ba...	netral	Mon Apr 03 12:59:15 +0000 2017
605	@AhokDjarot	2902333938	djarot blusukan pasar pagi nyonya kita2badja	netral	Mon Apr 03 12:45:02 +0000 2017
608	@AhokDjarot	2207200386	dukung warga madura jakarta kualitas integrita...	netral	Mon Apr 03 12:44:01 +0000 2017
609	@AhokDjarot	2902333938	rawa buaya djarot cerita ubah kawasan kalijodo...	netral	Mon Apr 03 12:44:01 +0000 2017
610	@AhokDjarot	822579060547158017	bidang sehat tingkat fasilitas rumah sakit pus...	netral	Mon Apr 03 12:43:18 +0000 2017
611	@AhokDjarot	2902333938	djarot kembang wisata kebun salak condet kita2...	netral	Mon Apr 03 12:43:01 +0000 2017
612	@AhokDjarot	822579060547158017	14 sisi rehab total puskesmas 34 lokasi camat ...	netral	Mon Apr 03 12:42:43 +0000 2017
616	@AhokDjarot	597598439	foto un ah ahok akun instagram nya sabtu 1 4 2...	netral	Mon Apr 03 12:38:01 +0000 2017
619	@AhokDjarot	820945995273904129	ahok bangun bioskop rakyat tiket rp10 ribu kit...	netral	Mon Apr 03 12:36:38 +0000 2017
621	@AhokDjarot	808168296860127233	36 program bukti kerja nyata bukti manf t war	netral	Mon Apr 03 12:36:01 +0000 2017
622	@AhokDjarot	597598439	hantam fitnah ahok rawan tangkis kita2badja	netral	Mon Apr 03 12:36:00 +0000 2017
625	@AhokDjarot	839351851824762880	ahok bangun bioskop rakyat tiket rp10 ribu kit...	netral	Mon Apr 03 12:34:12 +0000 2017

id	username	id user	isi	sentimen	Mon Apr 03 12:34:01 +0000 2017
626	@AhokDjarot	808168296860127233	34 bentuk pasu khusus biru oranye kuning hijau...	netral	
628	@AhokDjarot	597598439	bioskop rakyat sambut produser film eks kita2b...	netral	Mon Apr 03 12:33:01 +0000 2017
629	@AhokDjarot	808168296860127233	tata kota perhati serius era bersih indah amp ...	netral	Mon Apr 03 12:33:01 +0000 2017
630	@AhokDjarot	597598439	ahok bangun bioskop rakyat tiket rp10 ribu kit...	netral	Mon Apr 03 12:32:01 +0000 2017
631	@AhokDjarot	808168296860127233	32 wisata religi makam mbah priuk amp keliling...	netral	Mon Apr 03 12:32:01 +0000 2017
632	@AhokDjarot	822579060547158017	12 era rsud kualitas tara rumah sakit swasta n...	netral	Mon Apr 03 12:31:59 +0000 2017
634	@AhokDjarot	2207309154	tim sedia aplikasi khusus saksi tps kita2badja	netral	Mon Apr 03 12:31:01 +0000 2017
635	@AhokDjarot	812984602545373184	ahok bioskop lantai 3 4 pasar pasar jakarta ki...	netral	Mon Apr 03 12:31:01 +0000 2017
636	@AhokDjarot	597598439	ahok ide bioskop rakyat pasar pasar kita2badja	netral	Mon Apr 03 12:31:01 +0000 2017
640	@AhokDjarot	808168296860127233	30 bidang pariwisata bukti buka wisata kita2badja	netral	Mon Apr 03 12:30:01 +0000 2017
641	@AhokDjarot	597649776	rencana bangun bioskop pribadi milik perintah ...	netral	Mon Apr 03 12:30:01 +0000 2017
642	@AhokDjarot	810139401242214402	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
643	@AhokDjarot	777456541335035904	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
644	@AhokDjarot	809783453428264960	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
645	@AhokDjarot	821663582882500608	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
646	@AhokDjarot	810136877256863746	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
647	@AhokDjarot	821695975064412160	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017
648	@AhokDjarot	1270831718	tim gelar latih dukung apartemen kita2badja	netral	Mon Apr 03 12:29:59 +0000 2017

In [123]:



Word Cloud untuk Sentimen = 'negatif'

In [124]:

```
negative_data = data[data['sentimen'] == 'negatif']  
negative_data
```

Out[124]:

	akun	id_user	isi	sentimen	tanggal
2	@AhokDjarot	799108167170949121	golong sumbu pendek fpi piara keluarga cendana...	negatif	Tue Apr 04 13:56:37 +0000 2017
9	@AhokDjarot	1222354812	ahok penjara bersih2 penjara biar ahok	negatif	Tue Apr 04 13:39:44 +0000 2017
14	@AhokDjarot	230151878	wkwkwk udang dbalik batu	negatif	Tue Apr 04 13:34:27 +0000 2017
15	@AhokDjarot	230151878	dzolimi bgmana jela mendolimi umat islam sampe...	negatif	Tue Apr 04 13:33:24 +0000 2017
17	@AhokDjarot	230151878	nista agama y	negatif	Tue Apr 04 13:32:04 +0000 2017
69	@AhokDjarot	2207159918	jakarta bersyariah wujud jakarta berbhi eka pi...	negatif	Tue Apr 04 13:01:01 +0000 2017
78	@AhokDjarot	3314956485	dukung jujur bohong percaya	negatif	Tue Apr 04 12:04:47 +0000 2017
79	@AhokDjarot	1353959941	nya ya oke	negatif	Tue Apr 04 11:59:34 +0000 2017
80	@AhokDjarot	847636767490490368	bodoh dukung	negatif	Tue Apr 04 11:54:15 +0000 2017
81	@AhokDjarot	847636767490490368	dukung jujur bohong percaya	negatif	Tue Apr 04 11:53:19 +0000 2017
83	@AhokDjarot	845984159050825728	anis fokuswacana fokus kerja anis program kerja	negatif	Tue Apr 04 11:45:58 +0000 2017
85	@AhokDjarot	329545782	lempar handuk	negatif	Tue Apr 04 11:20:01 +0000 2017
87	@AhokDjarot	379496119	ibukota jakarta milik golongan milik jakartapuny...	negatif	Tue Apr 04 11:09:24 +0000 2017
88	@AhokDjarot	1587555818	hahaha lucu lomba ya	negatif	Tue Apr 04 11:04:23 +0000 2017
93	@AhokDjarot	821969963372007424	dukung jujur bohong percaya	negatif	Tue Apr 04 10:57:24 +0000 2017
98	@AhokDjarot	2913617202	dukung jujur bohong percaya	negatif	Tue Apr 04 10:53:59 +0000 2017
104	@AhokDjarot	2420918802	mas hitung2 sumbangsih nyata konglomerat rakya...	negatif	Tue Apr 04 10:19:11 +0000 2017
113	@AhokDjarot	791981359241715713	ush d tonton tayangin d	negatif	Tue Apr 04 09:52:19 +0000 2017
114	@AhokDjarot	733649502201118720	temu foto rencana kecurangan2 pasang waspada	negatif	Tue Apr 04 09:32:35 +0000 2017
115	@AhokDjarot	845522944289755136	pilih aza takabur h	negatif	Tue Apr 04 09:30:44 +0000 2017
116	@AhokDjarot	86898109	repost with dki jakarta cari kerja latar	negatif	Tue Apr 04 09:21:05 +0000 2017
118	@AhokDjarot	2420918802	this is informasi guys k p on blackcampaign wi...	negatif	Tue Apr 04 08:47:12 +0000 2017
119	@AhokDjarot	561817520	dapet sms sms nya	negatif	Tue Apr 04 08:42:57 +0000 2017
126	@AhokDjarot	3161593170	jpu kuasa hukum adu bukti sidang ahok menang	negatif	Tue Apr 04 08:22:06 +0000 2017
127	@AhokDjarot	2919157490	patunganrakyat buka yuk kampanyerakyat putar	negatif	Tue Apr 04 08:20:31 +0000 2017
129	@AhokDjarot	719667624	mesti bikin sinetron ahog ngepet segra masuk b...	negatif	Tue Apr 04 08:04:25 +0000 2017
130	@AhokDjarot	819558827582332928	lapor dk	negatif	Tue Apr 04 08:03:13 +0000 2017
131	@AhokDjarot	847246198658318336	muka2 firaun anti agama alias kominis	negatif	Tue Apr 04 07:58:51 +0000 2017
132	@AhokDjarot	836039925812215808	cek usaha in	negatif	Tue Apr 04 07:52:13 +0000 2017
135	@AhokDjarot	836039925812215808	bank pmliknya	negatif	Tue Apr 04 07:47:00 +0000 2017
...
554	@AhokDjarot	808251067922161664	kita2badja anies sandi hadir debat sandi ngamb...	negatif	Mon Apr 03 13:20:01 +0000 2017
556	@AhokDjarot	812984602545373184	hadir acara debat prabowo kecewa anies sandi p...	negatif	Mon Apr 03 13:19:02 +0000 2017
557	@AhokDjarot	808251067922161664	sabun al maidah aroma sara kita2badja	negatif	Mon Apr 03 13:17:01 +0000 2017
558	@AhokDjarot	809055759044481025	strategi perang kota ahok djarot makjleb bendu...	negatif	Mon Apr 03 13:16:00 +0000 2017
559	@AhokDjarot	76000792	nista politik p	negatif	Mon Apr 03 13:15:52 +0000 2017
560	@AhokDjarot	810495542983847937	rw basis kunjung rawan kuat pilih takut pilih	negatif	Mon Apr 03 13:15:01 +0000 2017


```
print(classifier.__class__.__name__)
print('****Results****')
print("Accuracy: %f (+/- %.2f)" % (scores.mean(), scores.std() * 2))
print("="*30 + "\n")
```

Klasifikasi dengan Fitur TF-IDF saja

In [169]:

```
classify(tf_idf_feature)
```

```
=====
KNeighborsClassifier
****Results****
Accuracy: 0.521446 (+/- 0.11)
=====

=====
DecisionTreeClassifier
****Results****
Accuracy: 0.570363 (+/- 0.14)
=====

=====
LinearSVC
****Results****
Accuracy: 0.559719 (+/- 0.17)
=====
```

Klasifikasi dengan Fitur Bag-of-Words saja

In [170]:

```
classify(bag_of_words_feature)
```

```
=====
KNeighborsClassifier
****Results****
Accuracy: 0.515158 (+/- 0.16)
=====

=====
DecisionTreeClassifier
****Results****
Accuracy: 0.565558 (+/- 0.14)
=====

=====
LinearSVC
****Results****
Accuracy: 0.581312 (+/- 0.14)
=====
```

Klasifikasi dengan Kedua Fitur

In [171]:

```
classify(np.concatenate((tf_idf_feature, bag_of_words_feature), axis=1))
```

```
=====
KNeighborsClassifier
****Results****
Accuracy: 0.500982 (+/- 0.16)
=====
```

```
=====
DecisionTreeClassifier
****Results****
Accuracy: 0.567306 (+/- 0.15)
=====

=====
LinearSVC
****Results****
Accuracy: 0.568956 (+/- 0.14)
=====
```

Analisis

Hasil analisis:

- Akurasi tertinggi didapatkan dengan menggunakan model **LinearSVC** dengan fitur Bag of Words saja.
- Sedangkan, akurasi terendah didapatkan dengan menggunakan model **KNeighbors** dengan penggabungan dua fitur tersebut.
- DecisionTreeClassifier merupakan classifier yang paling stabil dengan akurasi 56-57%
- KNeighborsClassifier merupakan classifier dengan nilai akurasi paling rendah di antara kedua classifier yang lain pada 3 fitur yang berbeda
- Penggunaan dua fitur secara sekaligus tidak meningkatkan secara drastis akurasi klasifikasi

In []:

In []: