

**PEMBUATAN DATASET KATA SINGKATAN BAHASA
INDONESIA MENGGUNAKAN METODE
CROWDSOURCING**

Skripsi



oleh

FRANSISKA LITMANTORO HIDAYAT 7115

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2019

**PEMBUATAN DATASET KATA SINGKATAN BAHASA
INDONESIA MENGGUNAKAN METODE
CROWDSOURCING**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi
Informasi Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam
Memperoleh Gelar Sarjana Komputer

Disusun oleh

FRANSISKA LITMANTORO HIDAYAT 71150166

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2019

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PEMBUATAN DATASET KATA SINGKATAN BAHASA INDONESIA MENGUNAKAN METODE CROWDSOURCING

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 17 Desember 2019



HALAMAN PERSETUJUAN

Judul Skripsi : PEMBUATAN DATASET KATA SINGKATAN
BAHASA INDONESIA MENGGUNAKAN
METODE CROWDSOURCING

Nama Mahasiswa : FRANSISKA LITMANTORO HIDAYAT

N I M : 71150166

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276

Semester : Gasal

Tahun Akademik : 2019/2020

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 17 Desember 2019

Dosen Pembimbing I



Danny Sebastian, S.Kom., M.M., M.T.

Dosen Pembimbing II



Kristian Adi Nugraha, S.Kom., M.T.

HALAMAN PENGESAHAN

**PEMBUATAN DATASET KATA SINGKATAN BAHASA INDONESIA
MENGUNAKAN METODE CROWDSOURCING**

Oleh: FRANSISKA LITMANTORO HIDAYAT / 71150166

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 13 Desember 2019

Yogyakarta, 17 Desember 2019
Mengesahkan,

Dewan Penguji:

1. Danny Sebastian, S.Kom., M.M., M.T.
2. Kristian Adi Nugraha, S.Kom., M.T.
3. Lukas Chrisantyo, S.Kom., M.Eng.
4. Junius Karel, S.Si., M.T.

Dekan DUTA WACANA

Ketua Program Studi


(Restyandito, S.Kom., MSIS, Ph.D.)


(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena telah melimpahkan rahmat dan kasih karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir/skripsi yang berjudul “Pembuatan Dataset Kata Singkatan Bahasa Indonesia Menggunakan Metode *Crowdsourcing*”.

Penulis menyusun skripsi dalam rangka untuk memenuhi salah satu persyaratan untuk mencapai gelar sarjana (S1) pada Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana Yogyakarta.

Dalam menyelesaikan program skripsi ini, penulis telah banyak mendapatkan bimbingan, saran, serta dukungan dari berbagai pihak. Maka dari itu, penulis mengucapkan terima kasih kepada:

1. Mama dan keluarga yang telah mendukung dan menemani penulis.
2. Bapak Danny Sebastian dan Bapak Kristian Adi Nugraha selaku dosen pembimbing yang telah membimbing selama penelitian.
3. Gabby, Clarissa, Steven Edgar, Kevin Valiant, Yadi, Dimas, Audri, Zenas, Albert dan teman-teman seperjuangan Jurusan Informatika UKDW terutama angkatan 2015 yang selalu ada untuk memberikan semangat dan berjuang bersama dalam menyelesaikan tugas akhir. Terakhir untuk setiap orang yang telah memberikan semangat kepada penulis namun tidak bisa dicantumkan satu per satu.

Dalam penyusunan tugas akhir ini, penulis masih memiliki banyak kekurangan. Oleh karena itu, penulis sangat menghargai dan meminta jika ada berbagai masukan dari para pembaca, baik berupa kritik maupun saran yang sifatnya membangun demi penyempurnaan penulisan-penulisan Skripsi dimasa yang akan datang. Akhir kata penulis ingin meminta maaf apabila terdapat kesalahan dalam penyusunan laporan maupun yang pernah penulis lakukan sewaktu pelaksanaan skripsi.

Yogyakarta, 20 Desember 2019

Penulis

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa atas segala berkah dan penyertaan-Nya, penulis dapat menyelesaikan skripsi berjudul “Pembuatan Dataset Kata Singkatan Bahasa Indonesia Menggunakan Metode *Crowdsourcing*” dengan lancar.

Dengan selesainya tugas akhir ini, tidak lepas dari bantuan banyak pihak yang telah memberikan masukan dan saran kepada penulis. Untuk itu penulis mengucapkan terima kasih.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kesempurnaan baik dari bentuk penyusunan maupun materinya. Oleh karena itu segala kritik dan saran yang membangun akan penulis terima dengan baik. Akhir kata semoga laporan tugas akhir ini dapat memberikan manfaat kepada kita sekalian.

Yogyakarta, 20 Desember 2019

Penulis

DAFTAR ISI

PERNYATAAN KEASLIAN SKRIPSI.....	Error! Bookmark not defined.
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASIError! Bookmark not defined.	
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
UCAPAN TERIMA KASIH	vi
KATA PENGANTAR.....	vii
INTISARI	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
1.6. Metode Penelitian.....	3
1.7. Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	6
1.1 Tinjauan Pustaka	6
1.2 Landasan Teori	8
1.2.1 Crowdsourcing	8
1.2.2 Dataset	9
1.2.3 Kata Singkatan	9
1.2.4 Twitter.....	11
1.2.5 IFTTT.....	12

1.2.6	<i>Text Mining</i>	13
1.2.7	<i>Text Preprocessing</i>	13
BAB 3	PERANCANGAN SISTEM	15
3.1.	Spesifikasi Sistem	15
3.1.1.	<i>Deskripsi Sistem</i>	15
3.1.2.	<i>Kebutuhan Sistem</i>	15
3.2.	Diagram Alir Sistem	16
3.2.1	<i>Flowchart Pengumpulan Data</i>	16
3.2.2	<i>Flowchart Pengisian Data Oleh Responden</i>	18
3.2.3	<i>Use Case</i>	19
3.3	Rancangan Database	20
3.4	Proses Pengolahan Data	23
3.5	Rancangan Antarmuka	25
3.6	Metode Pengujian	26
BAB 4	IMPLEMENTASI DAN ANALISIS SISTEM	28
4.1	Implementasi Sistem	28
4.2	Pengujian Sistem	32
4.2.1	<i>Analisis dan Pengujian Sistem</i>	32
BAB 5	KESIMPULAN DAN SARAN	41
5.1.	Kesimpulan	41
5.2.	Saran	41
DAFTAR PUSTAKA	42
LAMPIRAN	44

DAFTAR TABEL

Tabel 4. 1.....	34
Tabel 4. 2.....	35
Tabel 4. 3.....	35
Tabel 4. 4.....	36
Tabel 4. 5.....	36
Tabel 4. 6.....	37
Tabel 4. 7.....	37
Tabel 4. 8.....	37
Tabel 4. 9.....	38
Tabel 4. 10.....	39

©UKDW

DAFTAR GAMBAR

Gambar 2. 1. Ilustrasi <i>Crowdsourcing</i>	8
Gambar 2. 2. Logo Twitter.....	11
Gambar 2. 3. Tampilan applet pada Web IFTTT	12
Gambar 2. 4. tampilan data dari IFTTT ke Google Spreadsheet	13
Gambar 3. 1. Flowchart Pengambilan Data	16
Gambar 3. 2. Flowchart Pengisian Data Oleh Responden	18
Gambar 3. 3. Use Case Diagram	20
Gambar 3. 4. Rancangan Struktur Database	20
Gambar 3. 5. Isi Table Responden	21
Gambar 3. 6. Isi Tabel Responses	21
Gambar 3. 7. Isi Table response_tokens	22
Gambar 3. 8. Isi Tabel tweets.....	22
Gambar 3. 9. Isi Tabel summary	23
Gambar 3. 10. Contoh Kalimat Sumber Berisi Singkatan	23
Gambar 3. 11. Rancangan Mock Up Halaman Data Diri	25
Gambar 3. 12. Rancangan Mock Up Halaman Kuesioner	26
Gambar 4. 1. Antarmuka halaman utama.....	28
Gambar 4. 2. Antarmuka halaman petunjuk pengisian	29
Gambar 4. 3. Antarmuka halaman kuisisioner	30
Gambar 4. 4. Antarmuka petunjuk pengisian pada halaman kuisisioner.....	31
Gambar 4. 5. Antarmuka Halaman Penutup	31
Gambar 4. 6. Tampilan halaman cek hasil summary	32
Gambar 4. 7. Pie chart dari responden	33

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Bahasa merupakan sarana yang sangat penting untuk menyampaikan pesan, maksud, dan tujuan dalam berkomunikasi. Di Indonesia, Bahasa Indonesia merupakan media komunikasi utama yang digunakan oleh masyarakat Indonesia. Seiring dengan perkembangan zaman, penggunaan Bahasa Indonesia baik dalam kehidupan sehari-hari maupun dalam ruang lingkup media sosial, mulai menampakkan pergeseran ke arah yang lebih modern, khususnya ditandai dengan maraknya penggunaan abreviasi atau pemendekkan kata menjadi lebih singkat dalam berkomunikasi sehari-hari (Astuti, 2014).

Mengingat pentingnya peran media sosial dalam berkomunikasi di era sekarang ini, abreviasi pada Bahasa Indonesia pun berpengaruh seiring dengan generasi pengguna media sosial. Kridalaksana (2007:159) menjelaskan arti abreviasi sebagai proses penanggalan satu atau beberapa bagian leksem atau kombinasi leksem, sehingga jadilah bentuk baru yang berstatus kata. Abreviasi sendiri dibuat untuk memudahkan kebutuhan manusia untuk berbahasa secara praktis. Tentunya setiap generasi memiliki berbagai jenis abreviasi yang tidak sama. Dalam berkomunikasi antar generasi sering terjadi kebingungan dalam penggunaan abreviasi ini.

Untuk mengatasi masalah di atas, peneliti membentuk *dataset* yang berisi kata singkatan dalam Bahasa Indonesia dengan metode *crowdsourcing*. *Crowdsourcing* menurut Howe (2019) diartikan menjadi aktifitas atau tindakan yang dilakukan oleh suatu perusahaan atau institusi yang mengambil salah satu fungsi pekerjaan atau tugas yang seharusnya dilakukan oleh karyawannya menjadi disebarluaskan secara terbuka dan bebas untuk orang banyak atau kerumunan yang terkoneksi dengan jaringan komputer. Kata singkatan Bahasa Indone-

sia yang dikumpulkan didapat dari pengumpulan data melalui penyebaran kuesioner menggunakan *website* yang disebarakan kepada pengguna media sosial aktif.

Penelitian ini merupakan lanjutan dari penelitian Nugraha dan Sebastian (2018) yang berjudul “Pembentukan *Dataset* Topik Kata Bahasa Indonesia pada Twitter Menggunakan *TF-IDF & Constine Similarity*”, dimana perbedaannya terletak pada subjek dan isi *dataset* yang diteliti. Penelitian sebelumnya membentuk *dataset* topik, sedangkan pada penelitian ini membentuk *dataset* kata singkatan. Untuk subjek pada penelitian sebelumnya menggunakan akun-akun Twitter media elektronik nasional yang terverifikasi, seperti: Kompas dan Detik, sedangkan pada penelitian ini menggunakan akun-akun Twitter terverifikasi yang terdeteksi sering menggunakan abreviasi atau kata singkatan.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana cara membangun *dataset* dengan metode *crowdsourcing*?
2. Bagaimana tingkat akurasi pengujian *dataset* singkatan kata bahasa Indonesia?

1.3. Batasan Masalah

Dalam penelitian ini, penulis memberikan batasan masalah sebagai berikut:

1. Masukan berupa teks singkatan Bahasa Indonesia.
2. Sumber data berupa media sosial Twitter.
3. Responden merupakan pengguna media sosial dengan minimal pendidikan terakhir SMA yang aktif menggunakan media sosial.

Untuk data yang diambil juga memiliki batasan tersendiri, yaitu:

1. *Tweet* yang diambil dari akun yang sudah terverifikasi atau yang bercentang biru.
2. *Tweet* berisikan kalimat yang mengandung kata singkatan dalam Bahasa Indonesia.
3. *Tweet* dicuitkan dalam jangka waktu Februari sampai November 2019.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk membentuk *dataset* yang berisi kata singkatan Bahasa Indonesia.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah menyediakan *dataset* untuk kata singkatan Bahasa Indonesia untuk penelitian terkait selanjutnya.

1.6. Metode Penelitian

Dalam penelitian ini, penulis menggunakan beberapa metode, yaitu:

1. Studi Pustaka

Studi pustaka dilakukan dengan mempelajari literatur berupa buku, jurnal, dan tulisan ilmiah, yang memuat informasi berkaitan dengan *dataset*, kata singkatan, metode *crowdsourcing*, dan cara implementasinya ke dalam sistem.

2. Pengumpulan Data

Pengumpulan data uji dilakukan secara bertahap. Berikut ini tahap-tahap pengerjaannya:

- Tahap 1: Mengumpulkan akun-akun Twitter yang terverifikasi atau bercentang biru, dan sering mencuitkan kalimat yang mengandung kata singkatan atau abreviasi kata Bahasa Indonesia.
- Tahap 2: Mengumpulkan *tweet* menggunakan bantuan aplikasi IFTTT yang menghubungkan Twitter dan Google Spreadsheet dari akun-akun yang sudah terpilih dari tahap 1.
- Tahap 3: Memilih 150 *tweets* yang mengandung kata singkatan dari data yang sudah disimpan pada Google Spreadsheet.
- Tahap 4: Data yang sudah dipilih tadi siap dimasukkan ke *database* sebagai data uji.

3. Perancangan Sistem

Dalam tahapan ini, peneliti merancang antarmuka dan basis data yang akan digunakan oleh sistem yang dibuat. Data yang digunakan sebagai data uji adalah 150 *tweets* yang mengandung kata singkatan Bahasa Indonesia.

4. Pembangunan Sistem

Pembangunan sistem dilakukan menggunakan perangkat lunak Sublime Text 3, phpMyAdmin, MySQL, Twitter, IFTTT, Google Spreadsheet.

5. Pengujian Sistem

Pengujian sistem ini dilakukan setelah kuesioner disebarakan kepada responden secara acak dan memenuhi syarat yang telah ditentukan sebelumnya. Setelah mendapat jawaban dari responden, jawaban tersebut akan dihitung menggunakan perhitungan statistik. Dari hasil perhitungan itulah dapat dibentuk *dataset* kata singkatan Bahasa Indonesia.

1.7. Sistematika Penulisan

Sistematika laporan tugas akhir ini dibagi ke dalam 5 bab, yaitu:

Bab 1 Pendahuluan menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

Bab 2 Tinjauan Pustaka dan Landasan Teori menjelaskan penelitian-penelitian serupa yang sudah ada sebelumnya dan teori-teori dasar yang mendukung penelitian ini.

Bab 3 Perancangan Sistem menjelaskan tahapan dan penjelasan mengenai deskripsi, kebutuhan, proses, dan pengujian sistem.

Bab 4 Implementasi dan Analisis Sistem menjelaskan proses dalam pengerjaan atau penggunaan sistem pada *website*, *database*, serta pengecekan data.

Bab 5 Kesimpulan dan Saran yang berisi kesimpulan dari hasil penelitian ini, serta saran pengembangan sistem ke depannya.

©UKDW

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Pembangunan dataset kata singkatan Bahasa Indonesia menggunakan metode *crowdsourcing* dikatakan berhasil, walaupun terdapat kekurangan dari metode *crowdsourcing* yaitu tidak adanya pengawasan dalam pengisian kuesioner oleh responden sehingga terdapat responden yang mengisi kepanjangan (*full-form*) secara asal-asalan. Akibatnya 28% *sample* kepanjangan (*full-form*) yang diisi responden merupakan kepanjangan yang tidak sesuai dengan kepanjangan yang seharusnya. Adanya sistem *reward* berupa pulsa dan petunjuk pengisian pada sistem ini rupanya kurang membantu mengarahkan responden untuk mengisi kuesioner dengan baik dan benar.

Sistem dapat membangun dataset kata singkatan menggunakan data dari Twitter dan menghasilkan kata singkatan serta kepanjangan dari kata singkatan tersebut. Dari hasil perhitungan sampel data, didapatkan tingkat akurasi kebenaran dari kata singkatan Bahasa Indonesia pada dataset ini 72%, dari persentase ini dapat ditegaskan bahwa pembuatan dataset dikatakan berhasil.

5.2. Saran

Untuk mengatasi masalah diatas, disarankan menggunakan fitur KBBI online. Dengan adanya fitur KBBI online, kata baku (non-singkatan) yang sudah ada pada KBBI tidak akan dimunculkan untuk diedit oleh responden menjadi kata baku. Disarankan adanya perbaikan petunjuk yang diberikan agar lebih jelas serta sistem *reward* yang dinaikkan sehingga responden lebih serius mengisi kuesioner.

DAFTAR PUSTAKA

- Ana, H. (2015). *Akronim dalam Bahasa Indonesia Tinjauan Linguistik dan Sosio-Politis Perkembangannya*. FKIP Universitas Haluoleo
- Andriansyah, M., Oswari, T., & Prijanto, B. (2014). *Crowdsourcing: Konsep Sumber Daya Kerumunan dalam Abad Patisipasi Komunitas Internet*.
- Ardiyanto, A. (2017). *Klasifikasi Komentar Pada dataset Pemilu Presiden Indonesia 2014 Dengan Metode Improved K-nearest Neighbor*. (Undergraduate thesis, Duta Wacana Christian University, 2017). Retrieved from <http://sinta.ukdw.ac.id>
- Astuti, N. (2014). *Singkatan dan Akronim di Kalangan Remaja di Kota Bandung*, FPBS, UPI
- Connolly, T. M., & Beg, C. E. (2015). *Database systems: A practical approach to design, implementation, and management*. Boston: Pearson.
- Gebze, K. M. (2017). The positional accuracy of crowdsourced geographic data from open source web GIS. 2017 International Symposium on Geoinformatics (ISyG). doi:10.1109/isyg.2017.8280674
- Howe, J. (2009). *Crowdsourcing: why the power of the crowd is driving the future of business*. New York, NY: Three Rivers Press.
- Kridalaksana, H. (2007). *Pembentukan Kata dalam Bahasa Indonesia*. Jakarta: PT Gramedia Pustaka Utama.
- Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *Jurnal Sains Dan Seni ITS*, 6(2). doi: 10.12962/j23373520.v6i2.24419
- Kurniawan, J. (2017). *Klasifikasi Sentimen Terhadap dataset Komentar Pemilu Presiden 2014 Pada Facebook Menggunakan Support Vector Machine*.

(Undergraduate thesis, Duta Wacana Christian University, 2017). Retrieved from <http://sinta.ukdw.ac.id>

Nugraha, K.A., & Sebastian, D. (2018). Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity. 2018 4th Jurnal Teknik Informatika dan Sistem Informasi

Pole, A.S. (2018). *Perhitungan Tf-idf Pada dataset Sentipol*. (Undergraduate thesis, Duta Wacana Christian University, 2018). Retrieved from <http://sinta.ukdw.ac.id>

Pratama, A.K.I. (2017). Pembangunan Aplikasi Mobile *Crowdsourcing* untuk Pengumpulan Data. Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta

W., D. R. P. (2017). Perancangan Sistem Manajemen Akses database Sqlite Dengan Aes. (Undergraduate thesis, Duta Wacana Christian University, 2017). Retrieved from <http://sinta.ukdw.ac.id>

