

ANALISIS SENTIMEN PADA OPINI PENGGUNA MASKAPAI PENERBANGAN MENGUNAKAN HYBRID CUCKOO SEARCH

Nanang Fakhrrur Rozi¹, Fandi Arianto², Dian Puspita Hapsari³

^{1,2,3}Jurusan Teknik Informatika, Institut Teknologi Adhi Tama Surabaya
Email: ¹nanang@itats.ac.id, ²ozilarianto@gmail.com, ³dian.puspita@itats.ac.id

(Naskah masuk: 04 Desember 2018, diterima untuk diterbitkan: 13 Februari 2019)

Abstrak

Tingginya minat penggunaan pesawat terbang dipengaruhi oleh tingginya tingkat mobilitas masyarakat yang menuntut perpindahan kota dalam waktu yang singkat. Meski demikian, tidak semua maskapai penerbangan mampu memberikan layanan yang memuaskan bagi konsumennya. Kualitas layanan yang diberikan oleh suatu maskapai, baik dari segi keselamatan, keamanan, maupun kenyamanan, umumnya dapat diketahui melalui opini penumpang lainnya. Banyaknya opini negatif yang didapat oleh maskapai mengindikasikan buruknya kualitas layanannya, begitu pula sebaliknya. Akan tetapi, jumlah opini yang semakin hari semakin meningkat menyebabkan sulitnya konsumen dalam menilai kualitas maskapai secara cepat. Oleh karena itu, analisis sentimen dibutuhkan guna mempercepat konsumen dalam menilai kualitas layanan maskapai. Hybrid Cuckoo Search (HCS) merupakan salah satu metode yang dapat digunakan untuk melakukan analisis tersebut. Metode ini mampu mengelompokkan informasi secara cepat. Penelitian ini bertujuan untuk mengimplementasikan HCS dalam melakukan analisis sentimen pada data opini penumpang maskapai penerbangan. Hasil uji coba menunjukkan bahwa nilai rata-rata akurasi, *precision*, dan *recall* dari data opini 7 maskapai dengan 1.000 iterasi masing-masing sebesar 69,24%; 70,88%; dan 77,57%.

Kata kunci: analisis sentimen, Hybrid Cuckoo Search, K-Means, maskapai penerbangan, opini penumpang

SENTIMENT ANALYSIS ON PASSENGER OPINIONS AT AIRLINES COMPANY USING HYBRID CUCKOO SEARCH

Abstract

The high demands of airplanes usage are influenced by the increasing levels of people's mobility who want to trip from one city to another in a short time. However, not all airlines company could provide satisfactory services for the consumers. The quality of services provided by an airline, in terms of safety, security, and convenience, is usually known through passenger opinions. The number of negative opinions gained by airlines indicates its poor quality of service and vice versa. However, the increasing number of opinion increases the difficulty of the consumer in assessing the quality of the airline quickly. Therefore, sentiment analysis is needed to accelerate the consumer in assessing the quality of airline services. Hybrid Cuckoo Search (HCS) is a method which can be used in conducting such analysis. This method is able to group information quickly. This study aims to implement HCS in conducting sentiment analysis on airline passenger opinion data. The results show that the averaged accuracy, precision, and recall from opinion dataset of 7 airlines company at 1,000 iteration are 69.24%, 70.88%, and 77.57% respectively.

Keywords: sentiment analysis, Hybrid Cuckoo Search, K-Means, airlines company, passenger opinion

1. PENDAHULUAN

Pesawat terbang merupakan salah satu sarana transportasi udara yang banyak digunakan oleh masyarakat. Tingginya minat masyarakat untuk menggunakan pesawat terbang dipengaruhi oleh semakin tingginya tingkat mobilitas masyarakat yang menuntut untuk dapat berada di kota tertentu dalam waktu yang singkat. Hal tersebut sejalan dengan data statistik jumlah penumpang pesawat terbang di

Indonesia yang mencapai 94.722.823 penumpang (Oktober 2017) yang meningkat 6,1% dibanding tahun 2016 yang berjumlah 89.281.491 penumpang (Direktorat Jenderal Perhubungan Udara, 2017).

Kenaikan jumlah penumpang tersebut tidak serta-merta mengindikasikan tingkat keamanan dan keselamatan bagi penumpang. Hal itu antara lain dapat dibuktikan dengan insiden kecelakaan pesawat AirAsia Indonesia pada 28 Desember 2014 yang

mengakibatkan 162 penumpang meninggal dunia (Kompas.com, 2017).

Guna mengetahui kualitas layanan jasa maskapai penerbangan, baik dari segi keselamatan, keamanan, maupun kenyamanannya, para calon penumpang dapat memperhatikan opini penumpang sebelumnya. Saat ini, opini dan pengalaman yang ditulis oleh pengguna jasa layanan penerbangan secara daring terus meningkat. Opini tersebut banyak dituangkan pada forum-forum yang ada di internet, antara lain situs web yang bernama Skytrax.

Skytrax adalah perusahaan yang melakukan riset mengenai peringkat dan kualitas maskapai penerbangan sipil. Hasil riset tersebut didasarkan pada data situs web yang dikembangkannya. Pada situs tersebut terdapat kolom komentar yang dapat dimanfaatkan pengguna jasa maskapai penerbangan untuk memberikan ulasan mengenai kualitas dan layanan maskapai penerbangan yang ada di seluruh dunia sehingga dapat dijadikan acuan oleh calon penumpang lain dalam menilai kualitasnya.

Membaca ulasan tersebut secara keseluruhan tentu memakan waktu yang lama. Apabila hanya sedikit ulasan yang dibaca, evaluasi yang dihasilkan juga akan bias. Usaha untuk mengumpulkan semua ulasan tersebut dan mengolahnya menjadi informasi yang dapat menjawab keingintahuan para pengguna jasa sangat dibutuhkan untuk mengetahui kualitas maskapai yang nantinya akan dipilih. Usaha tersebut dapat dicapai dengan memanfaatkan teknik analisis sentimen.

Analisis sentimen, sering pula disebut sebagai penambangan opini (*opinion mining*), adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi orang-orang terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, atau topik tertentu (Liu, 2012). Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen. Dengan demikian, kecondongan sebuah teks untuk bersifat positif atau negatif dapat diketahui dengan mudah.

Teknik analisis sentimen yang pernah dilakukan antara lain teknik yang berbasiskan pembelajaran mesin tanpa supervisi (*clustering*). Pandey dkk. (2017) melakukan analisis sentimen menggunakan metode Hybrid Cuckoo Search (HCS) yang diimplementasikan pada 4 koleksi data yang diambil dari media sosial Twitter. HCS sendiri merupakan teknik pengelompokan data yang menggabungkan metode K-Means dengan Cuckoo Search. Penelitian tersebut mampu menghasilkan rata-rata akurasi sebesar 78%.

Pada penelitian ini, metode yang digunakan dalam analisis sentimen opini pengguna maskapai penerbangan juga HCS. Tahapan-tahapan dalam penelitian ini yaitu pemrosesan awal terhadap data, pengekstraksian fitur, pengelompokan data dengan HCS, dan pengevaluasian hasil.

2. METODE PENELITIAN

2.1 Pemrosesan Awal terhadap Data

Pemrosesan awal terhadap data merupakan tahap awal yang biasanya dilakukan dalam penambangan teks. Tahap ini mencakup semua rutinitas dan proses untuk menyiapkan data yang akan digunakan pada operasi penggalian informasi (Feldman & Sanger, 2007). Dalam penelitian ini, pemrosesan awal terhadap data dibagi menjadi dua fase. Fase pertama terdiri atas 5 proses yang berfungsi untuk meniadakan elemen derau (*noise*) pada setiap opini. Proses-proses tersebut antara lain:

- (1) menghilangkan URL dengan memanfaatkan Regular Expression;
- (2) menghilangkan @username;
- (3) menghapus # (tagar);
- (4) menghilangkan *forward slash (/)*, *backward slash (\)*, dan *dash (-)*; serta
- (5) menghapus spasi berlebih.

Fase kedua juga terdiri atas 5 proses yang berfungsi untuk menormalisasi teks opini lebih lanjut. Proses-proses tersebut memanfaatkan dua kamus yaitu kamus *stopword* dan kamus akronim. Adapun proses-proses yang termasuk dalam fase ini yaitu

- (1) mengubah semua kata ke dalam huruf kecil;
- (2) menghilangkan semua kata yang terdapat dalam kamus *stopword* seperti *as*, *a*, *is*, atau lainnya;
- (3) menghilangkan karakter yang berulang seperti karakter 'o' dalam *halooooo*;
- (4) menghilangkan kata yang tidak diawali dengan huruf alfabet; serta
- (5) mengubah semua singkatan menjadi bentuk kata penuh, seperti *gws* (*get well soon*).

2.2 Ekstraksi Fitur

Setelah melalui tahap pemrosesan awal terhadap data, setiap teks opini pengguna maskapai penerbangan akan diubah ke dalam bentuk vektor fitur yang terdiri atas 9 komponen fitur. Kesembilan komponen fitur tersebut yakni berupa hitungan (jumlah):

- (1) *kata*: mewakili jumlah kata yang tersedia di setiap teks opini;
- (2) *emoticon positif*: seperti *:*, *;*, *:D*, atau lainnya adalah simbol yang mengekspresikan momen bahagia (NetLingo, 2018);
- (3) *emoticon negatif*: simbol yang mengekspresikan momen sedih seperti *:* (atau lainnya (NetLingo, 2018);
- (4) *seruan positif*: kata seru, seperti *hore!*, *wow!*, atau lainnya, umumnya digunakan untuk menyampaikan perasaan atau pendapat yang sangat kuat tentang suatu topik;
- (5) *seruan negatif*: kata seruan (negatif), seperti *argh!* atau lainnya;
- (6) *penyangkalan*: mengekspresikan pendapat negatif, kata-kata penyangkalan seperti *no*, *not*, atau lainnya;

- (7) *kata positif*: kata positif berdasarkan kamus kata positif (Breen, 2015b);
- (8) *kata negatif*: kata negatif berdasarkan kamus kata negatif (Jeffreybreen, 2015a);
- (9) *kata intens*: kata-kata yang menggambarkan intensitas, seperti *very*, *much*, dan lain-lain., ditentukan berdasarkan kamus kata intens.

2.3 Pengelompokan dengan Hybrid Cuckoo Search

Metode Hybrid Cuckoo Search (HCS) merupakan metode yang menggabungkan K-Means dan Cuckoo Search. K-Means dimanfaatkan untuk menyelesaikan masalah inialisasi awal acak yang dilakukan oleh Cuckoo Search. Dalam metode ini, solusi yang diperoleh dari hasil pembangkitan *k* kelompok oleh K-Means digunakan untuk menginisialisasi populasi. Dengan demikian, ketika Cuckoo Search dijalankan, hasil yang optimal pun akan didapatkan. Selain itu, konvergensi juga akan diperoleh dengan waktu yang lebih cepat. Algoritme HCS menurut Pandey dkk. (2017) adalah sebagai berikut.

- (1) Berikan nilai *N* sebagai inialisasi banyaknya populasi.
- (2) For *i* to *N*
 - a) Jalankan K-Means untuk menghasilkan *k* kelompok.
 - b) Inialisasi populasi dari Cuckoo Search dengan *k* cluster hasil dari K-Means.
- (3) End For
- (4) Hitung nilai *fitness* dari *N* populasi menggunakan Persamaan (1).
- (5) While (*t* ≤ *MaxIterasi*)
 - a) Hasilkan *N* sebagai solusi baru dengan Levy Flight.
 - b) Hitung nilai *fitness* dari solusi baru tersebut dengan Persamaan (1).

$$\sigma^2 = \sum_{j=1}^N w_j (\mu_j - \mu)^2 \tag{1}$$

Pada Persamaan (1), *w_j* adalah probabilitas kejadian kelas *D_j*, *μ_j* adalah rata-rata dari kelas *D_j* dan *μ* adalah rata-rata total setiap fitur. Masing-masing *w_j*, *μ_j* dan *μ* dapat dicari menggunakan Persamaan (2–4).

$$w_j = \sum_{i \in D_j} p_i \tag{2}$$

$$\mu_j = \sum_{i \in D_j} i p_i / w_j \tag{3}$$

$$\mu = \sum_{i=1}^T i p_i \tag{4}$$

Pada Persamaan (4), *p_i* merupakan distribusi probabilitas masing-masing fitur yang dapat dicari menggunakan Persamaan (5).

$$p_i = \frac{O_i}{n} \tag{5}$$

Pada Persamaan (5), *i* mewakili nilai fitur *i* yang berada pada rentang $0 \leq i \leq T$ dan *O_i* menunjukkan jumlah data yang memiliki nilai fitur *i* dan *n* menunjukkan banyak data.

- c) Bandingkan solusi baru dengan solusi lama. Jika solusi baru lebih baik dari solusi lama maka ganti solusi lama dengan solusi yang baru.
 - d) Ganti *P_a* dari solusi terburuk dengan yang baru.
- (6) Cetak solusi terbaik.

2.4 Evaluasi Hasil

Evaluasi dilakukan dengan mengamati kinerja dari Hybrid Cuckoo Search yang meliputi tingkat akurasi, *precision*, dan *recall*. Koleksi data yang digunakan merupakan opini yang diambil dari situs web Skytrax. Terdapat empat fitur pada data tersebut, yakni nama maskapai, tanggal pengiriman (*posting*), teks opini, dan sentimen. Contoh datanya dapat dilihat pada Gambar 1 berikut.

vietnam-airlines
22/01/18
London to Hanoi. Vietnam Airlines lost my bag which was not a big deal but the handling of its recovery was the worst service I have ever received. They lost my bag on 5th January then found my bag 10 days later then couldn't find it again then it was being sent to Vietnam then I was told I would be compensated fully if 21 days passed as the bag would be declared lost finally after hours (no exaggeration) of trying to get service they acted and got me my bag after 16 days. \$30 is the height of monetary compensation which barely covers phone credit in the many attempts at trying to get through to their lost and found never mind clothes a new luggage bag etc. They rarely make contact and are impossible to get in contact with. They did not make contact after I logged my claim at 5am on the 5th January or 6th January and I had to call them on the 7th. When my bag was "found" in London I made contact 3 days later wondering where it was only to be told they did not know and it could be in Dublin.
NEGATIVE

Gambar 1. Contoh Data Opini

Dapat dilihat pada Gambar 1, salah satu penanda opini negatifnya didasarkan pada kata negatif *worst*.

Total data yang digunakan berjumlah 3.882 buah opini yang dipisahkan menjadi 7 koleksi data sesuai dengan nama maskapai penerbangan. Data yang diambil merupakan data pada kisaran April 2009 hingga Januari 2018. Rincian ketujuh koleksi data tersebut dapat dilihat pada Tabel 1.

Tabel 1. Rincian Koleksi Data

Nama Maskapai	Jml. Opini	Kelompok Sentimen	
		Positif	Negatif
AirAsia	498	306	192
Garuda Indonesia	623	571	52
Lion Air	109	41	68
Malaysia Airlines	900	569	331
Philippine Airlines	419	261	158
Singapore Airlines	855	675	180
Vietnam Airlines	478	289	189
Total	3.882	2.712	1.170

Tabel 1 menunjukkan koleksi data yang akan diujikan, meliputi data opini untuk maskapai AirAsia (498 opini), Garuda Indonesia (623 opini), Lion Air (109 opini), Malaysia Airlines (900 opini), Philippine Airlines (419 opini), Singapore Airlines (855 opini), dan Vietnam Airlines (478 opini). Uji coba dilakukan pada ketujuh koleksi data dengan jumlah iterasi yang bervariasi, yaitu 100, 300, 600, 800, dan 1.000. Pada akhirnya, nilai akurasi, *precision*, dan *recall* tiap pengujian atas koleksi data akan dirata-rata untuk memperoleh gambaran menyeluruh atas kinerja HCS.

3. HASIL DAN PEMBAHASAN

Hasil uji coba atas skenario pengujian yang telah dipaparkan pada subbab 1 dapat dilihat pada tabel-tabel yang akan dipaparkan pada subbagian berikut.

3.1 Uji Coba untuk AirAsia

Uji coba untuk data opini pengguna maskapai penerbangan AirAsia dapat dilihat pada Tabel 2.

Tabel 2. Kinerja HCS untuk Opini di AirAsia

Jml. Iterasi	Kinerja (%)		
	Akurasi	Precision	Recall
100	65,46	67,91	83,00
300	65,46	69,47	78,10
600	65,86	67,98	83,98
800	65,46	67,09	85,94
1.000	66,26	68,35	83,98

Tabel 2 menunjukkan bahwa akurasi terbaik terjadi saat jumlah iterasi 1.000, sedangkan *precision* terbaik saat jumlah iterasi 300, dan *recall* terbaik saat jumlah iterasi 800.

3.2 Uji Coba untuk Garuda Indonesia

Uji coba untuk data opini pengguna maskapai penerbangan Garuda Indonesia dapat dilihat pada Tabel 3.

Tabel 3. Kinerja HCS untuk Opini di Garuda Indonesia

Jml. Iterasi	Kinerja (%)		
	Akurasi	Precision	Recall
100	83,78	93,35	88,61
300	82,66	92,94	87,74
600	85,87	92,74	91,76
800	86,19	92,61	92,29
1.000	85,07	92,98	90,54

Tabel 3 menunjukkan bahwa nilai akurasi terbaik terjadi saat jumlah iterasi 800, sedangkan *precision* terbaik saat jumlah iterasi 100, dan *recall* terbaik saat jumlah iterasi 800.

3.3 Uji Coba untuk Lion Air

Uji coba untuk data opini pengguna maskapai penerbangan Lion Air dapat dilihat pada Tabel 4.

Tabel 4 menunjukkan bahwa nilai akurasi terbaik terjadi saat jumlah iterasi 800 dan 1.000, sedangkan

precision terbaik saat jumlah iterasi 800, dan *recall* terbaik saat jumlah iterasi 600.

Tabel 4. Kinerja HCS untuk Opini di Lion Air

Jml. Iterasi	Kinerja (%)		
	Akurasi	Precision	Recall
100	60,55	40,00	9,75
300	61,46	44,44	9,75
600	62,38	50,00	36,58
800	63,30	53,33	19,51
1.000	63,30	52,94	21,95

3.4 Uji Coba untuk Malaysia Airlines

Uji coba untuk data opini pengguna maskapai penerbangan Malaysia Airlines dapat dilihat pada Tabel 5.

Tabel 5. Kinerja HCS untuk Opini di Malaysia Airlines

Jml. Iterasi	Kinerja (%)		
	Akurasi	Precision	Recall
100	62,33	66,71	80,66
300	63,44	66,26	85,94
600	62,88	66,90	81,72
800	63,44	67,14	82,60
1.000	63,66	66,13	87,17

Tabel 5 menunjukkan bahwa nilai akurasi terbaik terjadi saat jumlah iterasi 1.000, sedangkan *precision* terbaik saat jumlah iterasi 800, dan *recall* terbaik saat jumlah iterasi 1.000.

3.5 Uji Coba untuk Philippine Airlines

Uji coba untuk data opini pengguna maskapai penerbangan Philippine Airlines dapat dilihat pada Tabel 6.

Tabel 6. Kinerja HCS untuk Opini di Philippine Airlines

Jml. Iterasi	Kinerja (%)		
	Akurasi	Precision	Recall
100	67,06	68,80	86,20
300	66,34	67,54	88,50
600	66,58	68,84	84,67
800	66,34	67,64	88,12
1.000	67,78	68,86	88,12

Tabel 6 menunjukkan bahwa nilai akurasi terbaik terjadi saat jumlah iterasi 1.000, sedangkan *precision* terbaik saat jumlah iterasi 1.000, dan *recall* terbaik saat jumlah iterasi 300.

3.6 Uji Coba untuk Singapore Airlines

Uji coba untuk data opini pengguna maskapai penerbangan Singapore Airlines dapat dilihat pada Tabel 7.

Tabel 7. Kinerja HCS untuk Opini di Singapore Airlines

Jml. Iterasi	Kinerja (%)		
	Akurasi	Precision	Recall
100	72,86	81,44	85,03
300	74,03	81,50	86,81
600	73,68	80,82	87,40
800	74,97	82,05	87,40
1.000	75,67	81,25	89,92

Tabel 7 menunjukkan bahwa nilai akurasi terbaik terjadi saat jumlah iterasi 1.000, sedangkan *precision* terbaik saat jumlah iterasi 800, dan *recall* terbaik saat jumlah iterasi 1.000.

3.7 Uji Coba untuk Vietnam Airlines

Uji coba untuk data opini pengguna maskapai penerbangan Vietnam Airlines dapat dilihat pada Tabel 8.

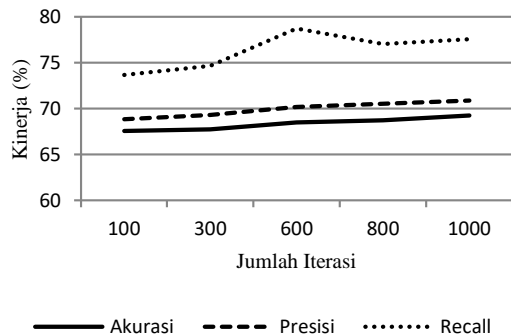
Tabel 8. Kinerja HCS untuk Opini di Vietnam Airlines

Jml. Iterasi	Kinerja (%)		
	Akurasi	Precision	Recall
100	60,87	63,63	82,35
300	61,08	63,10	85,81
600	62,34	64,22	85,12
800	61,71	64,09	83,39
1.000	62,97	65,64	81,31

Tabel 8 menunjukkan bahwa nilai akurasi terbaik terjadi saat jumlah iterasi 1.000, sedangkan *precision* terbaik saat jumlah iterasi 1.000, dan *recall* terbaik saat jumlah iterasi 300.

Hubungan Jumlah Iterasi dengan Kinerja HCS

Dengan merata-rata hasil uji coba tiap data opini maskapai penerbangan pada tiap-tiap titik iterasi yang dipilih, dapat dibuat grafik hubungan antara jumlah iterasi dalam HCS dengan kinerja hasilnya. Grafik tersebut dapat dilihat pada Gambar 2.



Gambar 2. Hubungan Jumlah Iterasi dengan Kinerja Metode

Gambar 2 menunjukkan bahwa peningkatan jumlah iterasi pada metode HCS secara umum berimplikasi pada peningkatan kinerja metode HCS dalam mengelompokkan sentimen opini penumpang maskapai, baik dilihat dari ukuran akurasi, *precision*, maupun *recall*. Nilai rata-rata yang bisa dicapai ketika jumlah iterasi 1.000 adalah 69,24% untuk akurasi, 70,88% untuk *precision*, dan 77,57% untuk *recall*.

4. SIMPULAN

Berdasarkan hasil uji coba yang telah dilakukan, HCS dapat digunakan sebagai salah satu metode untuk melakukan analisis sentimen atas opini penumpang maskapai penerbangan dengan koleksi data yang diambil dari situs web Skytrax. Selain itu,

jumlah iterasi pada HCS juga memengaruhi kinerja pengelompokan yang dilakukan.

DAFTAR PUSTAKA

- BREEN, J., 2015a. *Opinion Lexicon: Negative*. [online] Tersedia di: <https://github.com/jeffreymbreen/twitter-sentiment-analysis-tutorial-201107/blob/master/data/opinion-lexicon-English/negative-words.txt> [Diakses 5 Mar. 2018].
- BREEN, J., 2015b. *Opinion Lexicon: Positive*. [online] Tersedia di: <https://github.com/jeffreymbreen/twitter-sentiment-analysis-tutorial-201107/blob/master/data/opinion-lexicon-English/positive-words.txt> [Diakses 5 Mar. 2018].
- DIREKTORAT JENDERAL PERHUBUNGAN UDARA, 2017. *Tiga Tahun Tren Positif Capaian Penerbangan Nasional*. [online] Tersedia di: <http://hubud.dephub.go.id/?id/news/detail/3283> [Diakses 10 Des. 2017].
- FELDMAN, R. & SANGER, J., 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- KOMPAS.COM, 2015. Ini Analisis Lengkap Kecelakaan AirAsia QZ8501. *Kompas.com*. [online] Tersedia di: <http://tekno.kompas.com/read/2015/12/04/08090077/ini.analisis.lengkap.kecelakaan.airasia.qz8501> [Diakses 30 Jan. 2018].
- LIU, B., 2012. *Sentiment Analysis and Opinion Mining*.
- NETLINGO, 2018. *Smileys, Emoticons, Japanese Smileys, Assicons, Boobiecons, ASCII Art - NetLingo The Internet Dictionary: Online Dictionary of Computer and Internet Terms, Acronyms, Text Messaging, Smileys ;-)*. [online] Tersedia di: <https://www.netlingo.com/smileys.php> [Diakses 15 Jan. 2018].
- PANDEY, A.C., RAJPOOT, D.S. & SARASWAT, M., 2017. Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, 53(4), pp.764-779.

Halaman ini sengaja dikosongkan