

# ANALISIS SENTIMEN PASCA PEMILIHAN UMUM (PEMILU) INDONESIA TAHUN 2019 MELALUI PLATFORM TWITTER DENGAN *NAIVE BAYES*

---

**Audhi Aprilliant (G14160021)**

Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Institut Pertanian Bogor, Jl. Meranti Wing 22 Lantai 4,  
Dramaga, Bogor, Jawa Barat 16680

## PENDAHULUAN

### Latar Belakang

Menurut Solihah R (2018), pemilihan Umum 2019 adalah pemilihan legislatif dengan pemilihan presiden yang diadakan secara serentak. Hal ini dilakukan berdasarkan Putusan Mahkamah Konstitusi Nomor 14/ PUU-11/ 2013 tentang pemilu serentak, yang bertujuan untuk meminimalkan pembiayaan negara dalam pelaksanaan pemilu, meminimalisir politik biaya tinggi bagi peserta pemilu, serta politik uang yang melibatkan pemilih, penyalahgunaan kekuasaan atau mencegah politisasi birokrasi, dan merampingkan skema kerja pemerintah.

Pemilu Presiden 2019 mendatang diikuti oleh dua pasangan calon presiden dan wakil presiden. Pasangan nomor urut 01, Joko Widodo dan Ma'ruf Amin dicalonkan oleh PDI Perjuangan, Golkar, PKB, PPP, Nasdem, dan Hanura. Sementara, pasangan nomor urut 02, Prabowo Subianto dan Sandiaga Salahuddin Uno dicalonkan oleh Gerindra, PKS, PAN, dan Demokrat. Dua calon ini pernah bertarung pada pemilu presiden tahun 2014 lalu dengan selisih perolehan suara sebesar 6,3%. Saat itu, Joko Widodo mendapatkan suara sebesar 53,15%, dan Prabowo Subianto mendapatkan 46.85% suara. Dari 34 provinsi, pasangan Joko Widodo – M. Jusuf Kalla mengalami kekalahan di 10 provinsi yakni: Aceh, Sumatera Barat, Riau, Sumatera Selatan, Jawa Barat, Banten, Nusa Tenggara Barat, Kalimantan Selatan, Gorontalo, dan Maluku Utara (Fernandes A 2018).

Strategi kampanye di Pemilu Presiden 2019 tidak lagi menggunakan cara-cara yang konvensional, namun memanfaatkan teknologi digital seperti media sosial. Adanya perundang-undangan yang menjamin kebebasan individu untuk bebas mengeluarkan pendapat di muka umum, seperti media sosial menyebabkan kedua pendukung pasangan calon Jokowi-Ma'ruf Amin dan Prabowo Subianto-Sandiaga Salahudin Uno untuk mengeluarkan opini masing-masing terhadap pasangan calon presiden dan wakil presiden. Salah satu media sosial yang ramai digunakan oleh pendukung maupun masyarakat umum untuk mengeluarkan opininya mengenai Pemilu Presiden 2019 adalah Twitter. Menurut Ramdhani AL (2016), Twitter merupakan salah satu aplikasi jejaring dan media sosial yang sangat populer di Indonesia. Twitter di Indonesia banyak digunakan sebagai media penyebaran opini dan propaganda. Setiap pengguna Twitter menyampaikan opini dan propaganda dengan batasan 140 karakter disebut dengan *tweet*.

Oleh karena itu dibutuhkan suatu analisis deskriptif yang dapat digunakan untuk mendeskripsikan dan menggambarkan topik-topik yang dibahas oleh pendukung dan masyarakat Indonesia mengenai pasangan calon presiden dan wakil presiden Jokowi-Ma'ruf Amin dan Prabowo Subianto-Sandiaga Salahudin Uno. Dalam penelitian ini digunakan analisis sentimen dan visualisasi *wordcloud* sebagai metode analisis deskriptif.

## Tujuan

Tujuan dari penelitian mengenai Analisis Sentimen Pasca Pemilihan Umum (PEMILU) Indonesia Tahun 2019 melalui Platform Twitter adalah sebagai berikut: (1) Mengidentifikasi topik pembicaraan terhadap pasangan calon presiden Joko Widodo dan Prabowo Subianto pasca Pemilihan Umum 2019 di platform Twitter; (2) Membandingkan topik pembicaraan antara pasangan calon presiden Joko Widodo dan Prabowo Subianto pasca Pemilihan Umum 2019 di platform Twitter; (3) Melakukan analisis sentimen terhadap *tweet* pendukung dan masyarakat Indonesia terhadap pasangan calon presiden Joko Widodo dan Prabowo Subianto pasca Pemilihan Umum 2019 di platform Twitter; dan (4) Mengidentifikasi persebaran akun Twitter berdasarkan sentimen yang dihasilkan melalui *tweet* di platform Twitter terhadap pasangan calon presiden Joko Widodo dan Prabowo Subianto pasca Pemilihan Umum 2019 di platform Twitter.

## Manfaat

Manfaat yang dapat diperoleh dari penelitian mengenai Analisis Sentimen Pasca Pemilihan Umum (PEMILU) Indonesia Tahun 2019 melalui Platform Twitter adalah sebagai berikut: (1) Mengetahui topik pembicaraan berdasarkan *tweet* terhadap pasangan calon presiden Joko Widodo dan Prabowo Subianto pasca Pemilihan Umum 2019 di platform Twitter; (2) Mendapatkan gambaran mengenai sentimen antara terhadap pasangan calon presiden Joko Widodo dan Prabowo Subianto pasca Pemilihan Umum 2019 di platform Twitter; dan (3) Mengetahui persebaran akun Twitter berdasarkan sentimen yang dihasilkan melalui *tweet* di platform Twitter terhadap pasangan calon presiden Joko Widodo dan Prabowo Subianto pasca Pemilihan Umum 2019 di platform Twitter.

## Ruang Lingkup

Ruang lingkup pada penelitian mengenai Analisis Sentimen Pasca Pemilihan Umum (PEMILU) Indonesia Tahun 2019 melalui Platform Twitter adalah sebagai berikut: (1) Data yang digunakan pada penelitian ini merupakan data primer yang didapatkan melalui hasil *crawling* dari platform Twitter tanggal 28 – 29 Mei 2019; (2) Data dibagi menjadi dua bagian, data Joko Widodo sebanyak 1000 *tweet* yang memuat *keyword* “Joko Widodo” dan data Prabowo Subianto sebanyak 1000 *tweet* yang memuat *keyword* “Prabowo Subianto”; dan (3) Data bersifat acak dari *tweet* tanggal 28 – 29 Mei 2019 dengan batasan pada bahasa Indonesia.

## TINJAUAN PUSTAKA

### Analisis Sentimen

Analisis sentimen disebut juga sebagai klasifikasi sentimen. Menurut Medhata *et al.* (2014), analisis sentimen merupakan studi komputasi kata pada opini, sikap, dan emosi seseorang terhadap suatu entitas. Entitas tersebut dapat berupa orang, benda, kejadian atau suatu topik tertentu. Fungsi dasar dari analisis sentimen adalah mengklasifikasikan kecenderungan kata, kalimat atau dokumen ke dalam kelompok sentimen positif, negatif dan netral (Vinodhini & Chandrasekaran 2012).

Penelitian analisis sentimen berdasarkan Medhata *et al.* (2014) terdiri dari dua metode yaitu algoritme pembelajaran mesin dan kamus. Pertama metode algoritme pembelajaran mesin. Metode tersebut terdiri dari algoritme pembelajaran terbimbing (*supervised learning*) dan pembelajaran tidak terbimbing (*unsupervised learning*). Algoritme pembelajaran mesin yang digunakan seperti *Naïve Bayes* (Chinthala *et al.* 2015), *Support Vector Machines* (Mullen & Collier 2004), *Neural Network* (Ghiassi *et al.* 2013), dan *Artificial Neural Network* (Rodrigo *et al.* 2013).

Kedua metode kamus atau *lexicon*. Metode tersebut menggunakan kumpulan kata atau frasa serta metode statistik dan semantik untuk menentukan kecenderungan sentimen (Liu 2010). Sumber kamus yang dapat digunakan berasal dari bahasa Inggris di antaranya adalah *opinion lexicon* (Hu & Liu 2004), SentiWordNet (Esuli & Sebastiani 2006) dan AFINN (Nielsen 2011). Analisis sentimen dengan metode kamus pada data berbahasa Indonesia dilakukan oleh Lunando dan Purwarianti (2013) untuk mengidentifikasi sarkasme dengan menerjemahkan kamus SentiWordNet. Proses penerjemahan menggunakan Google Translate.

Selain kedua metode yang dijelaskan sebelumnya, terdapat pendekatan yang menggabungkan metode algoritme pembelajaran mesin dan kamus. Metode tersebut dinamakan *hybrid* (Medhata *et al.* 2014). Metode *hybrid* dilakukan oleh Denecke (2008), Khan *et al.* (2015) dan Khuc *et al.* (2012) untuk mendapatkan akurasi yang lebih baik dalam menentukan sentimen dokumen teks.

### **Pembobotan (*Term Weighting*)**

Pembobotan dilakukan apabila telah mendapatkan hasil *pre-processing*. Ketika melakukan pembobotan metode yang umum digunakan yaitu TF-IDF. Pengertian dari *term frequency* (TF) yaitu jumlah dari kemunculan kata di dokumen (Indriati & Ridok 2016 mengutip dari Feldman 2007). Pembobotan TF dapat dilihat pada persamaan di bawah ini.

$$W_{tf(t,d)} = 1 + \log TF(t, d)$$

Keterangan:

$W_{tf(t,d)}$  : bobot *term t* pada dokumen d

$TF(t, d)$  : jumlah munculnya *term t* di dokumen d

Inverse document frequency (IDF) merupakan suatu pembobotan berasal dari kandungan kata tertentu pada dokumen. Pembobotan TF-IDF dapat dilihat dalam persamaan di bawah ini.

$$W(t, d) = TF(t, d) * \log \frac{D}{Dt}$$

Keterangan:

$W(t, d)$  : bobot *term t* pada dokumen d

$TF(t, d)$  : jumlah munculnya *term t* di dokumen d

$D$  : jumlah keseluruhan dokumen

$Dt$  : banyaknya atau jumlah dokumen yang di dalamnya terkandung *term t*

Apabila proses TF-IDF telah dilakukan maka dilanjutkan dengan menghitung normalisasi (Nathania *et al.* 2018). Perhitungan normalisasi terdapat dalam persamaan di bawah ini.

$$W_{t,d} = \frac{W_{t,d}}{\sqrt{\sum_{t=1}^n W_{t,d}}}$$

### **Pengukuran Evaluasi**

Dalam menguji keefektifan suatu klasifikasi dibutuhkan suatu pengukuran evaluasi. Pengukuran tersebut didapatkan dalam sebuah set *confusion matrix*. *Confusion matrix* merupakan sebuah tabel yang terdiri atas banyaknya baris data uji yang diprediksi benar dan

tidak benar oleh model klasifikasi, digunakan untuk menentukan kinerja suatu model klasifikasi.

Tabel 1 *Confusion Matrix*

<i>Observation</i>		<i>Predicted</i>	
		<i>Positive</i>	<i>Negative</i>
	<i>Positive</i> <i>Negative</i>	TP FP	FN TN

Keterangan:

TP (*True Positive*) : kelas yang diprediksi positif dan benar

TN (*True Negative*) : kelas yang diprediksi negatif dan benar

FP (*False Positive*) : kelas yang diprediksi positif dan salah

FN (*False Negative*) : kelas yang diprediksi negatif dan salah

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}; \quad TP_{rate} = \frac{TP}{TP+FN}; \quad FP_{rate} = \frac{FP}{FP+TN}; \quad AUC = \frac{1+TP_{rate} - FP_{rate}}{2}$$

## METODE

### Data Penelitian

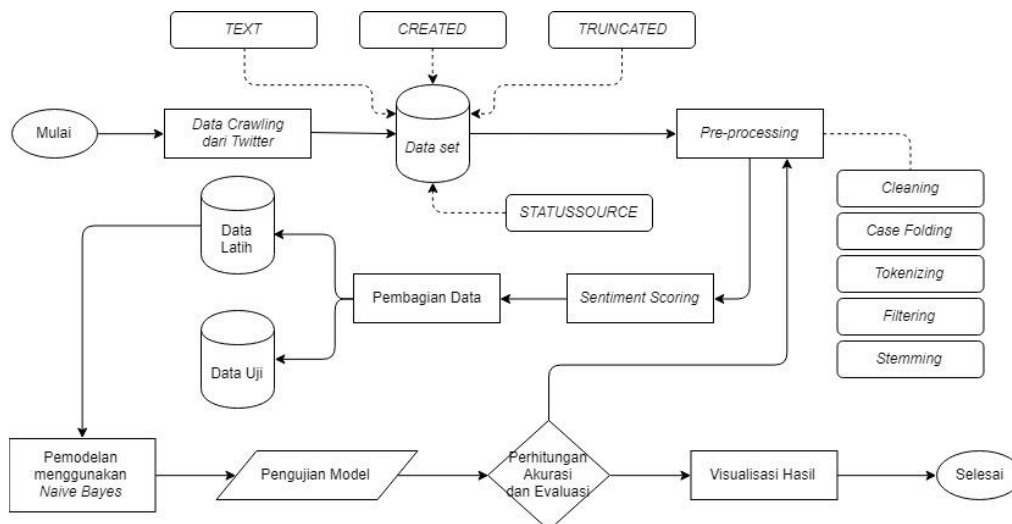
Data yang digunakan adalah data primer yang didapatkan melalui hasil *crawling* dari platform Twitter tanggal 28 – 29 Mei 2019 dengan format *csv (comma delimited)*. Data terbagi menjadi dua, yaitu data Joko Widodo yang memuat *keyword* “Joko Widodo” dan data Prabowo Subianto yang memuat *keyword* “Prabowo Subianto”. Data ini berisi beberapa peubah dalam menentukan sentimen *user*. Data ini terdiri dari 16 peubah atau atribut dan 1000 amatan atau user untuk masing-masing data. Beberapa peubah yang terdapat di dalam data ditampilkan pada tabel 2 sebagai berikut:

Tabel 2. Penjelasan Beberapa Peubah Data Penelitian

No.	Nama Peubah	Tipe	Penjelasan
1	TEXT	<i>Character</i>	<i>Tweet</i> dari <i>user</i> pengguna platform Twitter
2	CREATED	<i>Date</i>	Tanggal dan waktu <i>tweet</i> dibuat oleh <i>user</i>
3	TRUNCATED	<i>Logical</i>	Bernilai 1 apabila <i>tweet</i> dari <i>user</i> terpotong dan bernilai 0 apabila <i>tweet</i> dari <i>user</i> tidak terpotong
4	ID	<i>Factor</i>	ID <i>user</i> pengguna platform Twitter
5	STATUSSOURCE	<i>Factor</i>	Halaman URL untuk mengakses <i>tweet</i> secara online
6	SCREENNAME	<i>Factor</i>	Username <i>user</i> pengguna platform Twitter
7	RETWEETCOUNT	<i>Numeric</i>	Jumlah <i>retweet</i> dari <i>tweet</i> dari <i>user</i>
8	ISRETWEET	<i>Logical</i>	Bernilai 1 apabila <i>tweet</i> dari <i>user</i> merupakan hasil <i>retweet</i> dari <i>user</i> lain dan bernilai 0 apabila <i>tweet</i> dari <i>user</i> bukan merupakan hasil <i>retweet</i> dari <i>user</i> lain
9	RETWEETED	<i>Numeric</i>	Bernilai 1 apabila <i>tweet</i> dari <i>user</i> di- <i>retweet</i> oleh <i>user</i> lain dan bernilai 0 apabila <i>tweet</i> dari <i>user</i> tidak di- <i>retweet</i> oleh <i>user</i> lain

### Tahapan Kegiatan

Tahapan yang dilakukan pada penelitian mengenai Analisis Sentimen Pasca Pemilihan Umum (PEMILU) Indonesia Tahun 2019 melalui Platform Twitter terdiri atas *data crawling*, *data pre-processing (cleaning, case folding, data normalization, tokenizing, filtering, dan stemming)*, pembobotan kata menggunakan algoritma yang didapatkan dari *e-Commerce tokopedia*, dan melakukan visualisasi hasil analisis sentimen melalui *wordcloud* dan grafik lainnya yang menunjang. Secara sistematis, alur dari tahapan-tahapan kegiatan pada penelitian dapat dilihat pada gambar 1.



Gambar 2 Barplot *Total Tweets* dengan (a) *Keyword* “Joko Widodo” per Jam tanggal 29 Mei 2019 (b) *Keyword* “Prabowo Subianto” per Jam tanggal 28 - 29 Mei 2019

## Lingkungan Pengembangan

Perangkat keras yang digunakan dalam penelitian ini adalah komputer personal dengan spesifikasi sebagai berikut:

- Prosesor : AMD A8 – 7410
- Memory : 4GB
- VGA : Radeon (TM) R5 Graphics

Perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut:

- Sistem operasi Microsoft Windows Pro 10 (64-bit) dan Ubuntu Bionic Breaver
- Bahasa pemrograman R dengan packages *twitteR*, *tm*, *dplyr*, *stringr*, *katadasaR*, *wordcloud*, dan *ggplot2*
- Microsoft Excel 2016 sebagai media pengolah data tambahan, media penggabungan data, dan transformasi data

## HASIL DAN PEMBAHASAN

### Data Crawling

Terdapat tiga tahapan dalam melakukan *data crawling* pada platform Twitter. Tahapan pertama yang dilakukan adalah membuat akun Twitter dengan tujuan untuk memperoleh *API Key*, *API Secret*, *Access Token*, dan *Access Token Secret*. Tahapan kedua adalah *import library* atau *package twitteR* menggunakan Rstudio. Tahapan ketiga adalah menghubungkan aplikasi yang sedang dikembangkan dengan platform Twitter dengan cara memasukkan *API Key*, *API Secret*, *Access Token*, dan *Access Token Secret* pada bagian program.

Data yang digunakan pada penelitian ini adalah *tweet* yang memuat *keyword* “Joko Widodo” dan “Prabowo Subianto”. *Tweet* yang diambil masih berupa data mentah dan terdapat noise, seperti URL, *hashtag* (#), *username* (@), tanda baca, dan lain sebagainya. Berikut tabel 3 merupakan contoh penerapan *request* data terhadap *keyword* “Joko Widodo” dan “Prabowo Subianto”.

Tabel 3. Hasil *Data Crawling* pada Peubah TEXT

No.	Keyword	TEXT
1	Joko Widodo	RT @randyhartono: https://t.co/NwXFKgB4Kw https://t.co/7GEIHWU0vt Petisi cabut kewarganegaraan Briziek @disruptivesigna @DasarKita

No.	Keyword	TEXT
2	Joko Widodo	#Emak2Anak2JadiTumbal Presiden Joko Widodo (Jokowi) mengecam pelaku teror yang melibatkan anak-anak dalam melancarkan aksinya.
3	Prabowo Subianto	RT @do_ra_dong: Sebenarnya yang justru dikuatirkan adalah keselamatan Prabowo Subianto dan timnya Sudah dikriminalisasi, handphonenya dik...
4	Prabowo Subianto	RT @tvOneNews: Sejumlah orang yang menamakan dirinya sebagai aktivis 98 melaporkan Capres 02 Prabowo Subianto dan sejumlah tokoh lain ke Ba...

### Data Pre-processing

Tahapan setelah proses *crawling* dari platform Twitter dalam penelitian mengenai Analisis Sentimen Pasca Pemilihan Umum (PEMILU) Indonesia Tahun 2019 melalui Platform Twitter adalah tahap pra-proses. Pra-proses yang dilakukan dapat dijabarkan sebagai berikut:

1. *Cleaning*: digunakan untuk menghapus karakter yang dianggap kurang penting seperti username (@), hashtag (#), URL, tanda baca (*punctuation*), angka, dan *emoticon* (Nathania *et al.* 2018 mengutip dari Nur & Santika 2011)
2. *Case folding*: digunakan dalam pengubahan huruf kapital menjadi kecil (Puspitasari *et al.* 2018)
3. *Data normalization*: digunakan untuk melakukan perbaikan pada kata-kata yang mengalami kesalahan penulisan atau *typo*.
4. *Tokenizing*: digunakan dalam pemisahan kata menggunakan *whitespace*. Pada *tokenizing* terdapat istilah *term* yang berarti kata (Amin FF 2012).
5. *Filtering*: digunakan untuk menghapus kata yang termasuk dalam daftar *stopwords*. Kata yang dihapus merupakan kata yang dianggap tidak relevan atau tidak berarti (Amin FF 2012).
6. *Stemming*: digunakan untuk menghapus imbuhan pada kata (Amin FF 2012 mengutip dari Tala 2003). Pada penelitian ini menggunakan algoritma dari Nazief-Adriani dalam menyelesaikan proses *stemming*. Tahapan pada *stemming* ini yaitu sebagai berikut (Imbar *et al.* 2014):
  - 6.1 Pengecekan kata yang belum melewati proses *stemming* yang terdapat pada kamus. Apabila ada, maka algoritma berhenti
  - 6.2 Menghapus *inflection suffixes* (“-lah”, “-ku”, “-nya”, dan “-mu”). Namun apabila terdapat (“-tah”, “-lah”, “-pun”, atau “-kah”), maka proses akan diulang yang berguna untuk penghapusan *possesive pronouns* (“-mu”, “-nya”, “-ku”) pada kamus
  - 6.3 *Derivation suffixes* dihapus (“-kan”, “-i”, “-kan”). Namun algoritma akan berhenti ketika kata masuk dalam kamus dan apabila sebaliknya, maka dilakukan langkah selanjutnya yaitu:
    - 6.3.1 Apabila “-an” telah dihapus dan dibagian akhir kata terdapat huruf “k”, maka “-k” dihilangkan. Namun apabila kata tersebut ada pada kamus, maka algoritma berhenti, sedangkan apabila sebaliknya, maka akan dilakukan langkah selanjutnya
    - 6.3.2 Akhiran dari kata yang sebelumnya telah melewati tahap penghapusan (“-an”, “-kan”, atau “-i”) akan dibawa kembali
  - 6.4 *Derivational prefix* dihapus (“be-“, “me“, “di-“, “te-“, “ke-“, “pe-“, atau “se-“). Apabila pada kamus terdapat kata tersebut, maka algoritma akan berhenti. Apabila sebaliknya, maka dibutuhkan *recoding*. Apabila selanjutnya terdapat proses seperti yang dijabarkan di bawah ini, maka algoritma berhenti
    - 6.4.1 Adanya kombinasi pada kata berupa awalan dan akhiran yang tidak diizinkan

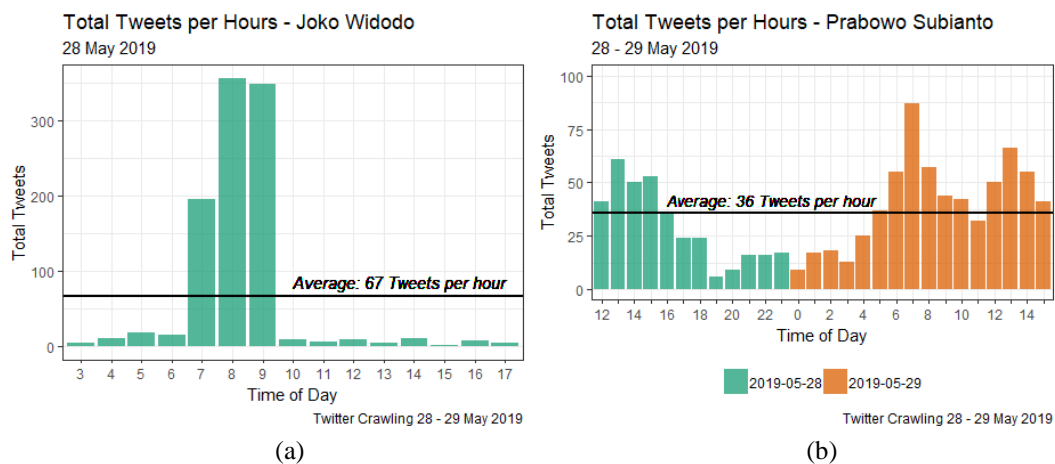
- 6.4.2 Terdapat awalan pada kata yang sama persis seperti yang telah dihapuskan sebelumnya
- 6.4.3 Terdapat adanya tiga awalan yang dihapuskan
- 6.5 Apabila langkah-langkah tersebut sudah dilewati, namun kata tersebut tetap tidak termasuk dalam kamus, maka sebelum proses *stemming* dilakukan kata tersebut akan dikembalikan

Tabel 4. Hasil *Pre-Processing* pada Peubah TEXT

No.	Keyword	TEXT
1	Joko Widodo	rt petisi cabut kewarganegaraan briziek
2	Joko Widodo	emak anak jaditumbalpresiden joko widodo jokowi kecamat teror libat anak anak lancar aksi
3	Prabowo Subianto	rt dikuatirkan selamat prabowo subianto tim kriminalisasi handphoneya dik
4	Prabowo Subianto	rt orang nama aktivis lapor capres prabowo subianto tokoh ba

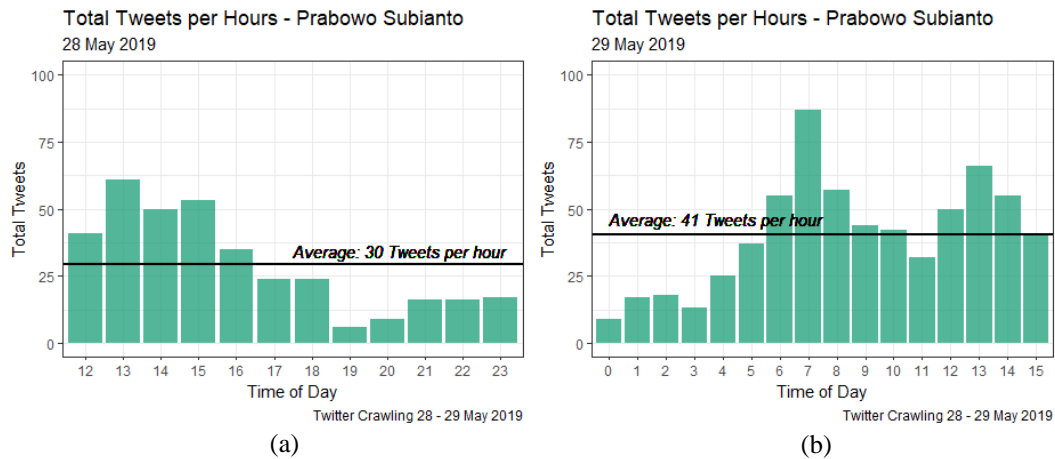
### Eksplorasi Data

Eksplorasi data digunakan untuk mendapatkan informasi dari data tweet hasil *data crawling* yang telah dilakukan *pre-processing*. Eksplorasi dilakukan untuk beberapa peubah yang dianggap cukup menarik untuk dibahas. Peubah tersebut seperti CREATED.



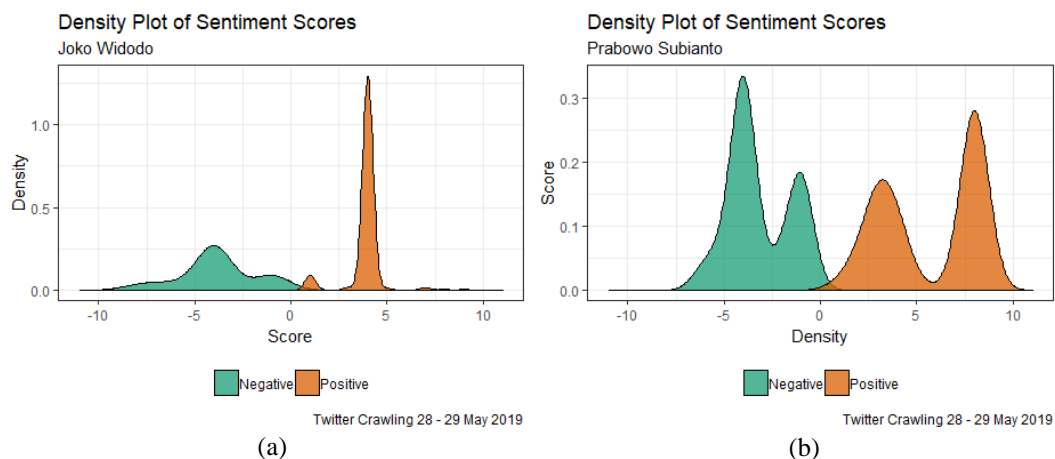
Gambar 2 Barplot *Total Tweets* dengan (a) *Keyword* “Joko Widodo” per Jam tanggal 29 Mei 2019 (b) *Keyword* “Prabowo Subianto” per Jam tanggal 28 - 29 Mei 2019

Berdasarkan gambar 2 dapat disimpulkan bahwa proses *data crawling* untuk *tweets* yang memuat *keyword* “Joko Widodo” dan “Prabowo Subianto” tidak persis menghasilkan *tweets* yang diunggah di tanggal yang sama. Sebagai contoh di gambar 2(a) dapat disimpulkan bahwa untuk *tweets* dengan *keyword* “Joko Widodo” hanya diperoleh di tanggal 28 Mei 2019 pada periode jam 03.00 hingga 17.00. Sedangkan pada gambar 2(b) dapat disimpulkan bahwa untuk *tweets* dengan *keyword* “Prabowo Subianto” diperoleh di tanggal 28 – 29 Mei 2019 dengan periode jam 12.00 hingga 23.59 di tanggal 28 Mei 2019 dan periode jam 00.00 hingga 15.00 di tanggal 29 Mei 2019. Berdasarkan gambar 2 diperoleh perbedaan yang signifikan antara pengguna platform Twitter yang mengunggah *tweet* dengan *keyword* “Joko Widodo” dengan *keyword* “Prabowo Subianto”. *Tweets* dengan *keyword* “Joko Widodo” cenderung sangat intens membicarakan Joko Widodo pada jam-jam tertentu (07.00 hingga 09.00) dengan jumlah *tweet* terbanyak pada jam 08.00 sebanyak 348 *tweets*. Sedangkan untuk *tweets* dengan *keyword* “Prabowo Subianto” cenderung konstan membicarakan Prabowo Subianto pada jam-jam di tanggal 28 – 29 Mei 2019. Rata-rata *tweets* dengan *keyword* “Prabowo Subianto” per jam yang diunggah pada tanggal 28 – 29 Mei 2019 adalah 36 *tweets* per jam.



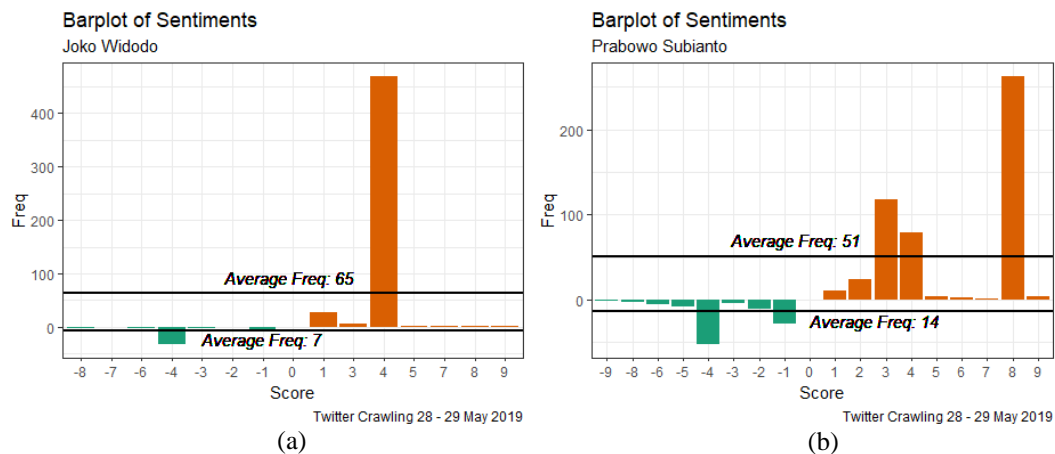
Gambar 3 Barplot *Total Tweets* dengan Keyword “Prabowo Subianto” per Jam  
(a) tanggal 28 Mei 2019 dan (b) tanggal 28 Mei 2019

Gambar 3 merupakan *barplot* dari jumlah *tweets* dengan keyword “Prabowo Subianto” tanggal 28 – 29 Mei 2019. Berdasarkan gambar 3(a) dapat disimpulkan bahwa rata-rata *tweets* yang diunggah oleh pengguna Twitter per jam mengenai Prabowo Subianto cenderung lebih sedikit dibandingkan dengan total *tweets* dalam dua hari sebesar 30 *tweets* per jam. Pengguna Twitter cenderung tidak aktif membicarakan topik Prabowo Subianto mulai jam 19.00 hingga 23.59. Hal ini dikarenakan merupakan jam istirahat masyarakat Indonesia. Namun demikian, sebenarnya *tweets* mengenai topik ini tidak berhenti di jam-jam malam dikarenakan adanya kemungkinan pengguna dari luar negeri yang memiliki zona waktu berbeda dengan Indonesia maupun masyarakat Indonesia yang sedang beraktivitas pada jam malam. Pengguna mulai aktif kembali membicarakan topik Prabowo Subianto mulai jam 04.00 pagi hingga pada puncaknya di jam 07.00, kemudian turun hingga naik kembali pada jam 12.00. Hal ini menunjukkan kebiasaan pengguna menggunakan platform Twitter pada jam sebelum bekerja atau beraktivitas dan waktu istirahat siang.



Gambar 4 *Density Plot* dari *Sentiment Score* tweets tanggal 28 – 29 Mei 2019 dengan  
Keyword (a) “Joko Widodo” dan (b) “Prabowo Subianto”

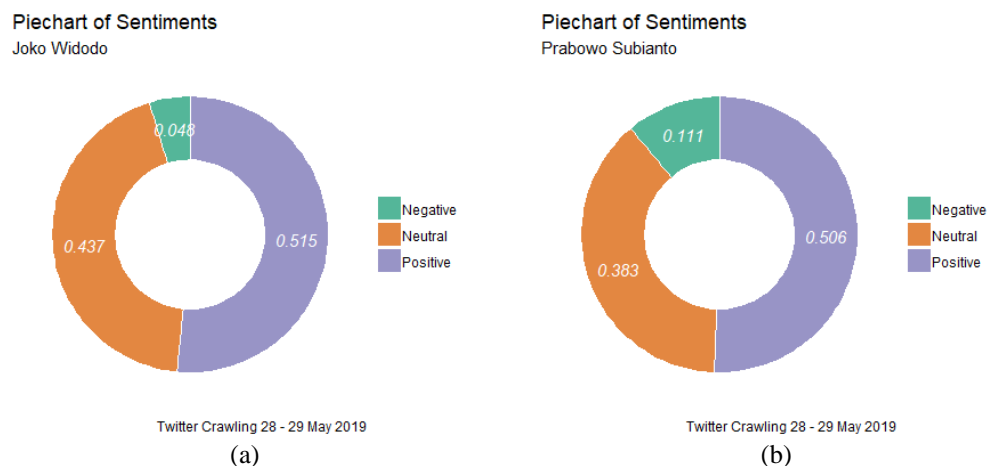




Gambar 5 Barplot dari *Sentiment Score* tweets tanggal 28 – 29 Mei 2019 dengan *Keyword* (a) “Joko Widodo” dan (b) “Prabowo Subianto”

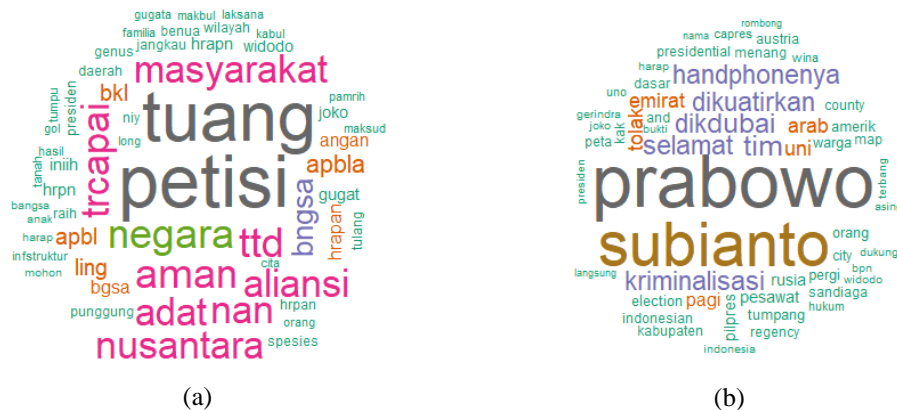
Gambar 4 merupakan *density plot* dari *sentiment score tweets* yang memuat *keyword* “Joko Widodo” dan “Prabowo Subianto”. *Score* untuk setiap *tweets* didapatkan dari rata-rata *score* pada tiap kata dasar penyusun *tweets* tersebut. Sehingga *score* diberikan untuk tiap kata dasar. *Sentiment score* memiliki interval dari -10 hingga 10. Semakin kecil *score*, maka sentimen untuk kata di dalam *tweets* tersebut semakin negatif. Demikian sebaliknya. Berdasarkan gambar 4(a) didapatkan kesimpulan bahwa sentimen negatif untuk *tweets* yang memuat *keyword* “Joko Widodo” berkisar dari -10 hingga -1 dengan median *score* di kisaran -4. Sentimen positif juga berkisar dari 1 hingga 10 dengan median di kisaran 4. Berdasarkan *density plot* di gambar 4(a) didapatkan bahwa *score* untuk sentimen positif memiliki keragaman yang cukup kecil. Sehingga disimpulkan bahwa sentimen positif untuk *tweets* dengan *keyword* “Joko Widodo” cenderung seragam.

Gambar 4(b) menunjukkan *density plot* dari *sentiment score tweets* yang memuat *keyword* “Prabowo Subianto”. Berbeda dengan gambar 4(a), sentimen negatif di gambar 4(b) berkisar di -8 hingga -1. Hal ini menunjukkan bahwa *tweets* tidak memiliki sentimen negatif yang terlalu besar (*tweets* memiliki sentimen yang negatif, tetapi tidak cukup besar). Selain itu, sebaran dari *score* sentimen negatif memiliki 2 puncak di kisaran angka 4 dan 1. Sedangkan untuk sentimen positif berkisar dari 1 hingga 10. Berbeda dengan gambar 4(a), sentimen positif di gambar 4(b) memiliki keragaman yang cukup besar dan memiliki 2 puncak di kisaran *score* 3 dan 10. Hal ini menunjukkan bahwa *tweets* yang memuat *keyword* “Prabowo Subianto” memiliki sentimen yang positif dengan nilai yang cukup besar.



Gambar 6 Pie Chart dari *Sentiment Class* tweets tanggal 28 – 29 Mei 2019 dengan *Keyword* (a) “Joko Widodo” dan (b) “Prabowo Subianto”

Gambar 6 merupakan ringkasan dari *sentiment score* tiap *tweets* yang telah dilakukan kategorisasi menjadi sentimen negatif, netral, dan positif. Sentimen negatif merupakan sentimen yang memiliki nilai *score* kurang dari nol. Sentimen netral adalah sentimen yang memiliki nilai *score* sama dengan nol. Sedangkan sentimen positif merupakan sentimen yang memiliki nilai *score* lebih dari nol. Berdasarkan gambar 6 didapatkan bahwa *tweets* yang memuat *keyword* “Joko Widodo” memiliki persentase jumlah sentimen negatif yang lebih kecil dibandingkan dengan *tweets* yang memuat *keyword* “Prabowo Subianto” dengan perbedaan sebesar 6,3%. Selain itu juga didapatkan bahwa *tweets* yang memuat *keyword* “Joko Widodo” memiliki persentase sentimen netral dan sentimen positif yang lebih besar dibandingkan dengan persentase sentimen netral dan sentimen positif *tweets* yang memuat *keyword* “Prabowo Subianto”. Secara eksplorasi melalui *piechart* didapatkan bahwa *tweets* dengan *keyword* “Joko Widodo” cenderung memiliki persentase sentimen positif yang lebih besar dibandingkan dengan *tweets* dengan *keyword* “Prabowo Subianto”, namun melalui eksplorasi *density plot* didapatkan bahwa sebaran dari *score* sentimen positif dan *score* sentimen negatif menunjukkan bahwa *tweets* dengan *keyword* “Prabowo Subianto” cenderung memiliki *sentiment score* lebih besar dibandingkan dengan *tweets* dengan *keyword* “Joko Widodo”. Sehingga perlu dilakukan pengujian lebih lanjut.



Gambar 7 Wordcloud Total Tweets dengan *Keyword* “Prabowo Subianto” per Jam  
(a) tanggal 28 Mei 2019 dan (b) tanggal 28 Mei 2019

Gambar 7 menunjukkan kata-kata di dalam *tweet* yang sering diunggah oleh pengguna platform Twitter yang memuat *keyword* “Joko Widodo” dan “Prabowo Subianto” pada tanggal 28 – 29 Mei 2019 (*data crawling*). Melalui visualisasi *wordcloud* ini dapat ditemukan topik-topik hangat yang menjadi bahan pembicaraan untuk masing-masing *keyword*. Untuk *tweets* yang memuat *keyword* “Joko Widodo” didapatkan bahwa kata “tuang”, “petisi”, “negara”, “aman”, dan “nusantara” merupakan lima kata dengan jumlah kemunculan yang tertinggi di setiap *tweets*. Sedangkan Untuk *tweets* yang memuat *keyword* “Joko Widodo” didapatkan bahwa kata “Prabowo”, “Subianto”, “kriminalisasi”, “selamat”, dan “dubai” merupakan lima kata dengan jumlah kemunculan yang tertinggi di setiap *tweets*. Hal ini secara tidak langsung menunjukkan pola dari *tweet* yang diunggah dengan *keyword* “Prabowo Subianto” yaitu setiap *tweet* yang diunggah hampir dipastikan memuat nama Prabowo Subianto secara langsung, bukan melalui *mention* (@). Hal ini dikarenakan pada *pre-processing*, *mention* (@) akan dihapus.

### Text Classification

Menurut Jurka T (2012), beberapa metode *supervised text classification* seperti: *Supervised Latent Dirichlet Allocation* (SLDA), *Support Vector Machine*, *Multinomial Logistic Regression (maximum entropy)*, *Naive Bayes*, *Neural Network*, *Decision Tree*, *Random forest*, *Boosting*, dan *Bagging algorithm*. Pada penelitian ini akan digunakan algoritma *Naive Bayes*. Label yang digunakan adalah sentimen dari *tweets* yang memuat *keyword* “Joko Widodo” dan “Prabowo Subianto” yaitu *Positive*, *Negative*, dan *Neutral*.

Tabel 5. Metrik Evaluasi Klasifikasi *Naive Bayes* dengan *10-Fold Cross Validation* Data Joko Widodo

No.	Data	Accuracy	Positive			Negative			Neutral		
			Sen	Spe	GM	Sen	Spe	GM	Sen	Spe	GM
1	Data latih	0.88	0.89	1.00	0.94	1.00	0.90	0.94	0.86	0.95	0.90
2	Data uji	0.85	0.82	1.00	0.90	0.92	0.88	0.89	0.87	0.94	0.90

Tabel 6. Metrik Evaluasi Klasifikasi *Naive Bayes* dengan *10-Fold Cross Validation* Data Prabowo S

No.	Data	Accuracy	Positive			Negative			Neutral		
			Sen	Spe	GM	Sen	Spe	GM	Sen	Spe	GM
1	Data latih	0.87	0.87	0.98	0.92	0.87	0.90	0.88	0.86	0.93	0.89
2	Data uji	0.80	0.81	0.96	0.88	0.60	0.88	0.72	0.84	0.86	0.84

Keterangan:

Sen : Sensitivity

Spe : Specificity

GM : G-Mean

Berdasarkan tabel 5 dan 6 dapat disimpulkan bahwa *text classification* menggunakan algoritma *Naive Bayes* memiliki metrik evaluasi *accuracy*, *sensitivity*, *specificity*, dan *G-Mean* yang cukup tinggi (di atas 0.80). Dikarenakan klasifikasi yang dilakukan menggunakan 3 kelas sentimen, yaitu *Positive*, *Negative*, dan *Neutral*, maka pemeriksaan metrik evaluasi dilakukan untuk setiap kelas. Algoritma *Naive Bayes* dianggap kurang mampu mengklasifikasikan dengan baik kelas *Negative* pada data *tweets* yang memuat *keyword* “Prabowo Subianto”. Hal ini dapat dilihat dari metrik evaluasi *sensitivity* yang bernilai 0.60 (di bawah 0.80) sehingga menurunkan nilai *G-Mean*. Namun secara keseluruhan, akurasi yang didapatkan sudah cukup baik. Data *tweets* yang memuat *keyword* “Joko Widodo” maupun “Prabowo Subianto” tidak mengalami *overfitting*. Hal ini ditunjukkan oleh *accuracy* dari data latih dan data uji yang tidak jauh berbeda dengan *accuracy* data latih lebih besar dibandingkan data uji.

Berdasarkan penelitian yang telah dilakukan oleh Susilowati E *et al.* (2015) mengenai Implementasi Metode *Support Vector Machine* untuk melakukan Klasifikasi Kemacetan Lalu Lintas pada Twitter didapatkan beberapa hasil mengenai pengaruh komposisi data pada hasil klasifikasi yaitu sebagai berikut:

- Akurasi algoritma klasifikasi dipengaruhi oleh jumlah data yang digunakan. Rata rata dari hasil pengujian yang dilakukan semakin banyak data yang digunakan, maka hasil akurasi yang didapatkan juga semakin tinggi. Hal ini dikarenakan dengan semakin banyak data yang digunakan, maka semakin banyak keberagaman atau jenis data yang dapat ditangani oleh sistem sehingga ketika melakukan klasifikasi secara otomatis model mampu menangani berbagai macam pola
- Akurasi algoritma klasifikasi dipengaruhi oleh komposisi data latih dan data uji. Semakin banyak jumlah data yang diberikan pada data latih, maka semakin tinggi akurasi yang didapatkan. Hal tersebut terjadi karena dengan banyaknya data latih, model yang terbentuk dapat menangani lebih banyak keberagaman data yang ada pada data sehingga mampu mengklasifikasikan dengan lebih baik ketika melakukan pengujian

Berdasarkan penelitian yang telah dilakukan oleh Astuti Y (2015) mengenai Klasifikasi Posting Twitter Cuaca Provisi DIY menggunakan Algoritma C4.5 untuk Informasi pada Web Pariwisata didapatkan hasil mengenai *running time pre-processing* yaitu semakin besar jumlah sampel yang digunakan, maka semakin besar rata-rata *tweet* yang dapat diolah pada proses *pre-processing* dan TF-IDF dalam waktu satu detik.

## **SIMPULAN DAN REKOMENDASI**

### **Simpulan**

Simpulan yang didapatkan berdasarkan penelitian adalah bahwa (1) Sentimen negatif dari *tweets* yang memuat *keyword* “Prabowo Subianto” lebih tinggi dibandingkan dengan *tweets* yang memuat *keyword* “Joko Widodo”, namun apabila diperhatikan lebih seksama didapatkan rata-rata nilai sentimen positif untuk *tweets* yang memuat *keyword* “Prabowo Subianto” lebih tinggi dibandingkan dengan *tweets* yang memuat *keyword* “Joko Widodo” dan (2) *Text classification* dengan menggunakan algoritma *Naive Bayes* diperoleh metrik evaluasi *accuracy*, *sensitivity*, *specificity*, dan *G-Mean* yang cukup tinggi (di atas 0.80) dan tidak terdapat gejala *overfitting*, baik pada data *tweets* yang memuat *keyword* “Prabowo Subianto” maupun yang memuat *keyword* “Joko Widodo” sehingga dapat dijadikan sebagai gambaran awal dalam membandingkan algoritma *text classification* lainnya.

### **Rekomendasi**

Rekomendasi yang dapat diberikan kepada pihak yang hendak melakukan penelitian mengenai analisis sentimen pada platform Twitter menggunakan bahasa Indonesia adalah: (1) Proses *stemming* dapat dioptimalkan dengan menggunakan algoritma *stemming* bahasa Indonesia lain yang dianggap lebih baik; (2) Menambahkan daftar kesalahan penulisan, bahasa alay, dan singkatan ke dalam kamus *lexicon* agar mendapatkan hasil yang baik dalam proses *data normalization*; dan menggunakan metode *text classification* lain, seperti *Support Vector Machine*, *Multinomial Logistic Regression (maximum entropy)*, *Naive Bayes*, *Neural Network*, *Decision Tree*, *Random forest*, *Boosting*, atau *Bagging algorithm* untuk membandingkan kinerja dari masing-masing algoritma dalam melakukan klasifikasi.

## DAFTAR PUSTAKA

- Amin FF. 2012. Sistem Temu Kembali Informasi dengan Metode Vector Space Model [internet]. [diakses pada 2019 Juni 02]. Tersedia pada [http://ejournal.undip.ac.id/index.php/jsi\\_nbis](http://ejournal.undip.ac.id/index.php/jsi_nbis).
- Astuti Y. 2015. Klasifikasi Posting Twitter Cuaca Provisi DIY menggunakan Algoritma C4.5 untuk Informasi pada Web Pariwisata. *Makalah*. Dalam: Seminar Nasional Sains dan Teknologi Fakultas Teknik Universitas Muhammadiyah Jakarta, 17 November.
- Fernandes A. 2018. Politik identitas dalam Pemilu 2019: proyeksi dan efektivitas. *Center for Strategic and International Studies*. 1(1): 1 – 10.
- Medhta W, Hassan A, Korashy H. 2014. Sentiment analysis algorithm and applications: a survey. *ASEJ*. 5(1): 1093 – 1113.
- Puspitasari AA, Sntoso E, Indriti. 2018. Klasifikasi dokumen tumbuhan obat menggunakan metode improved k-nearest neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 2(2): 486 – 492.
- Ramdni AL. 2016. *Pemilihan Akun Berpengaruh pada Data Twitter menggunakan Skyline Query dalam Mapreduce Framework* [tesis]. Bogor (ID): Sekolah Pascasarjana Institut Pertanian Bogor.
- Solihah R. 2018. Peluang dan tantangan Pemilu serentak 2019 dalam perspektif politik. *Jurnal Ilmiah Ilmu Pemerintahan (JIIP)*. 3(1): 73 – 88.
- Susilowati E, Sabariah MK, Gozali AA. 2015. Implementasi metode *support vector machine* untuk melakukan klasifikasi kemacetan lalu lintas pada twitter. *E-Proceeding of Engineering*. 2(1): 1 – 7.
- Vinodhini G, Chandrasekaran RM. 2012. Sentiment analysis and opinion mining: a survey. *IJARCSSE*. 2(1): 283 – 292.