

# Analisis Sentimen Terhadap Pemilihan Presiden Indonesia 2019 Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes

Lukito Agung Waskito, Kemas Muslim Lhaksana,  
Danang Triantoro Murdiansyah

Fakultas Informatika, Universitas Telkom, Bandung  
lukitoagung@students.telkomuniversity.ac.id, kemasmuslim@telkomuniversity.ac.id,  
danangtri@telkomuniversity.ac.id,  
Prodi S1 Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

---

## Abstrak

Pada tahun 2019 ini, media sosial twitter masih menjadi media sosial yang digemari oleh banyak orang, walaupun jumlah pengguna aktif yang sempat mengalami penurunan yang signifikan akibat ragamnya media sosial lainnya. Pada media sosial twitter ini kita dapat menuangkan apapun yang ada di pikiran kita dalam bentuk gambar, suara, dan tulisan. Setiap hari banyak sekali hal-hal yang bisa dibahas dan kita bisa menuangkannya pada media sosial twitter. Mulai dari musik, komedi, ataupun mengenai politik dan lain sebagainya. Dari hal-hal yang bisa dibahas tersebut kita bisa menggali informasi yang bermanfaat dalam sebuah penelitian mengenai topik tertentu. Untuk mengolah data dari opini-opini dalam media sosial twitter tersebut kita bisa mengerjakannya dengan teknik analisis sentimen atau opinion mining. Tetapi dalam melakukan analisis sentimen tersebut, kita perlu menggunakan teknik analisis yang tepat agar informasi yang nantinya kita dapatkan bisa maksimal dan dapat bermanfaat. Oleh karena itu, pada makalah tugas akhir ini dilakukan sebuah analisis sentimen terhadap berita yang ada di media sosial twitter mengenai pemilihan umum presiden dan wakil presiden menggunakan metode *naïve Bayes classifier* dengan mengklasifikasikan sentimen menjadi positif, dan negatif. Hasil penelitian yang telah dilakukan memperoleh tingkat akurasi sebesar 71.67%. Hal tersebut menunjukkan bahwa analisis sentimen menggunakan metode *naïve Bayes* memperoleh hasil yang cukup baik. Hasil analisis sentimen ini dapat digunakan untuk melihat bagaimana masyarakat Indonesia khususnya pada media sosial twitter dalam menanggapi proses rangkaian pemilihan umum presiden Indonesia 2019.

Kata kunci : pilpres, twitter, media sosial, *naïve Bayes*, klasifikasi.

---

## Abstract

In 2019, social media twitter is still a popular social media for many people, even though the number of active users has decreased significantly due to the variety of other social media. On this social media twitter we can pour whatever is on our mind in the form of images, sounds, and writing. Every day there are lots of things that can be discussed and we can put them on social media twitter. Starting from music, comedy, or about politics and so forth. From things that can be discussed, we can explore useful information in a study on a particular topic. To process data from opinions on Twitter social media we can do it with sentiment analysis techniques or opinion mining. But in conducting this sentiment analysis, we need to use the right analytical techniques so that the information we get later can be maximized and can be useful. Therefore, in this final paper a sentiment analysis of the news on social media twitter regarding the election of the president and vice president using the *naïve Bayes classifier* method by classifying sentiments into positive, and negative. The results of the research conducted have an accuracy rate of 71.67%. This shows that sentiment analysis using the *naïve Bayes* method has good results. The results of this sentiment analysis can be used to see how Indonesian society especially on social media twitter in response to the process of the series of 2019 Indonesian presidential elections.

Keyword : twitter, presidential election, social media, *naïve Bayes*, classification

---

## 1. Pendahuluan

### 1.1 Latar belakang

Pada era globalisasi seperti sekarang ini, akses untuk mendapatkan sebuah informasi sangat mudah. Internet dari waktu ke waktu menunjukkan signifikansi yang terlihat jelas. Hal tersebut juga dipicu dengan semakin banyak perangkat gawai yang beredar di pasaran, disamping itu perangkat gawai tersebut mempunyai ragam biaya, dan keunggulannya masing-masing, bahkan dengan harga yang terjangkau kita bisa mendapatkan perangkat gawai yang mempunyai fitur-fitur yang bisa menunjang kebutuhan kita dalam mengakses sebuah informasi. Biaya akses internet

juga semakin murah dibandingkan dengan beberapa tahun kebelakang yang mengakibatkan jumlah pengguna internet di Indonesia semakin bertambah pesat. Seperti hasil riset dari *Wearesosial Hootsuite* yang penulis kutip dari *databoks.katadata.co.id*, yang dirilis Januari 2019 pengguna media sosial di Indonesia mencapai 150 juta atau sebesar 56% dari total populasi [1]. Jumlah tersebut naik 20% dari survei sebelumnya. Sementara pengguna media sosial pada media gawai mencapai 130 juta atau sekitar 48% dari populasi. Media sosial adalah sebuah media untuk berkomunikasi, bersosialisasi satu sama lain tanpa dibatasi oleh waktu, dan tempat. Media sosial mempunyai ragam dan fungsi yang berbeda-beda, salah satu media sosial yang populer adalah twitter. Pada media sosial twitter kita dapat berkomunikasi dengan orang lain, berbagi sebuah informasi berupa gambar, suara, dan tulisan. Setiap harinya pada twitter memiliki pembahasan mengenai banyak hal, mulai dari hobi, konsultasi dokter, politik dan lain sebagainya.

Dari sekian banyak hal yang dibahas di twitter setiap harinya tersebut kita bisa memanfaatkannya untuk mendapatkan sebuah informasi yang bermanfaat, dan hal tersebut dinamakan dengan analisis sentimen [2]. Opini-opini yang dilakukan dalam penelitian ini, dikelompokkan pada dua atribut yaitu positif dan negatif [7]. Pada penelitian ini, dibahas mengenai analisis sentimen terhadap pemilihan umum presiden dan wakil presiden yang sering muncul pada media sosial twitter. Salah satu metode yang digunakan untuk mengerjakan analisis sentimen adalah *naïve Bayes classifier*. Algoritma *naïve Bayes* merupakan sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *naïve Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi / kejadian. Keuntungan penggunaan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

## 1.2 Tujuan

Adapun Tujuan dari Tugas Akhir ini adalah sebagai berikut :

- Melihat bagaimana sentimen analisis yang ada di masyarakat dari media sosial twitter mengenai pemilihan umum presiden Indonesia 2019.
- Menerapkan metode *naïve Bayes Classifier* pada data twitter yang di gunakan.
- Mengetahui tingkat akurasi, presisi dan recall pada metode *naïve Bayes* dan hasil dari perhitungan yang didapatkan.

## 2. Studi Terkait

### 2.1 Media Sosial Twitter

Jejaring sosial merupakan tempat dimana orang-orang atau organisasi berkomunikasi satu sama lain tanpa terhalang oleh ruang dan waktu [10]. Sejarah dari jejaring sosial juga cukup panjang, mulai dari tahun 1997 dengan munculnya *sixdegrees.com*, *livejournal*, dan lain-lain. Lambat laun semakin beragam jenis dari jejaring sosial tersebut. Para developer jejaring sosial berlomba-lomba membuatnya dengan banyak fitur-fitur andalan masing-masing. Beberapa jejaring sosial yang populer adalah *Friendster*, *facebook*, *instagram*, *twitter*. Setiap dari jejaring sosial ini mempunyai ciri khas dan peminatnya masing-masing. Munculnya beragam media sosial ini juga membuat banyak akses untuk banyak hal semakin mudah, bahkan perusahaan-perusahaan banyak yang meminta orang (individu) ataupun organisasi kecil untuk menjadi pengiklan dari produk yang mereka buat.

Walaupun kita tahu bahwa media sosial tersebut mempunyai manfaat yang sangat banyak, tetapi media sosial juga mempunyai dampak negatif. Dampak negatif yang muncul dari media sosial adalah munculnya berita-berita bohong yang dikelola oleh orang-orang yang tidak bertanggung jawab, munculnya penipuan-penipuan yang tentunya membuat rugi beberapa orang, dan beberapa dampak negatif lainnya. Disamping itu salah satu dampak positif dari media sosial twitter ini adalah adanya informasi yang bisa digali dan informasi tersebut bisa bermanfaat [9]. Seperti contohnya dari tugas akhir ini, tugas akhir ini mengambil informasi dari media sosial twitter mengenai pemilu Indonesia 2019. Informasi tersebut bisa digunakan untuk melihat analisis sentimen masyarakat Indonesia mengenai rangkaian proses pemilu 2019.

### 2.2 Pemilihan umum Presiden dan Wakil Presiden Indonesia 2019

Pemilihan umum Presiden dan Wakil Presiden adalah proses demokrasi pada suatu wilayah guna menentukan Presiden dan Wakil Presiden pada masa jabatan tertentu. Pemilu di Indonesia diselenggarakan oleh Komisi Pemilihan Umum. Pemilu mempunyai beberapa tahapan dalam pelaksanaannya, mulai dari perencanaan program dan anggaran,

penyusunan peraturan KPU, sosialisasi, pendaftaran serta verifikasi peserta pemilu, laporan dan audit dana kampanye, masa tenang, pemungutan perhitungan suara rekapitulasi perhitungan suara, penetapan hasil pemilu presiden dan wakil presiden, dan terakhir adalah pengucapan sumpah janji.

Masa penyelenggaraan kegiatan semua di atas dilakukan dari tahun 2017 hingga 2019. Dari beberapa kegiatan masa pemilu yang ditunggu oleh masyarakat Indonesia salah satunya adalah debat calon Presiden dan Wakil Presiden. Untuk debat ini sendiri dilakukan secara beragam, tidak hanya untuk debat calon presiden dan calon wakil presiden secara terus menerus dalam lima kali jadwal debat tersebut. Debat pemilihan presiden ini terdiri dari debat antara calon presiden dan calon wakil presiden, debat antara calon presiden, debat antara calon wakil presiden, dan debat antara calon presiden. Untuk isu yang akan dibahas adalah sebagai berikut. Untuk debat pertama, tema akan seputar isu hukum di Indonesia. Tema debat kedua mengacu seputar isu seperti sumber daya alam, energi dan pangan, lingkungan hidup, serta infrastruktur. Tema debat ketiga berkaitan dengan isu kesejahteraan. Tema debat keempat fokus pada isu pemerintahan. Seperti ideologi, pemerintah, pertahanan dan keamanan serta internasional. Sedangkan tema kelima yakni isu ekonomi dan kesejahteraan sosial, keuangan dan investasi, perdagangan.

Rangkaian Debat pilihan presiden yang dikoordinir oleh Komisi Pemilihan Umum bertujuan untuk memberikan info dan gambaran bagaimana visi, misi, dan pandangan dari calon presiden dan wakil presiden jika nantinya terpilih. Pada masa debat ini masyarakat akan disuguhkan dengan paparan dari masing-masing calon, dan pada media sosial seperti twitter, instagram, facebook, whatsapp dan lain sebagainya banyak masyarakat Indonesia yang melakukan diskusi-diskusi terhadap hasil debat yang dilakukan oleh calon. Dan dari data-data media sosial inilah yang nantinya diambil untuk dilihat bagaimana informasi yang akan diperoleh mengenai pemilu ini.

### 2.3 Algoritma Naïve Bayes

*Naïve Bayes Classifier* merupakan sebuah metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan Teorema *Bayes* dengan asumsi ketidaktergantungan (independen) yang tinggi. Keuntungan penggunaan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yg diperlukan dalam proses pengklasifikasian [6]. Penggunaan metode *naïve Bayes Classifier* pada penelitian ini didasarkan pada banyaknya dataset yang dipakai sehingga membutuhkan suatu metode yang mempunyai performansi yang cepat dalam pengklasifikasian serta keakuratan yang cukup tinggi. Keuntungan penggunaan *naïve Bayes Classifier* adalah metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

Untuk menghitung probabilitas sentimen menggunakan rumus dibawah ini :

$$V_{MAP} = \underset{V_j \in V}{\text{arg max}} \prod_{i=1}^n P(x_i | V_j) P(V_j) \tag{1.1}$$

Sedangkan untuk menggunakan *Laplace smoothing* pada naïve Bayes menggunakan rumus di bawah ini :

$$P(t_k | c) = \frac{w_{ct} + 1}{(\sum_{w' \in v} W'_{ct}) + B'} \tag{1.2}$$

$w_{ct}$  = Pembobotan kata

$\sum_{w' \in v} W'_{ct}$  = Jumlah total W dari keseluruhan kata yang berada di kelas c.

$B'$  = Jumlah total W kata unik di semua kelas.

Contoh perhitungan metode naïve Bayes berdasarkan data pemilihan umum Presiden Indonesia 2019 :  
 Contoh data pemilihan umum Presiden Indonesia 2019 bisa dilihat pada tabel 1 berikut :

Tabel 1. Contoh Data

Label	Tweet
+	Pemilu bersih aman damai
-	Bagaimana pemilu damai kalau cebong kampret saling hina
+	Menuju pemilu prabowo presiden saya
+	Debat jokowi saya bangga
-	Saya tidak ada niat untuk ikut andil dalam pemilu
?	Pemilu dapat berjalan damai

Keterangan :

- + = data bernilai positif
- - = data bernilai negatif
- ? = data yang belum mempunyai nilai/label
- Data diatas mempunyai data training berjumlah 5, dan data testing berjumlah 1.

- Data training yang berjumlah 5 tersebut dipisahkan sesuai nilai atau labelnya yaitu positif dan negatif [8].
- Setelah dipisahkan sesuai labelnya, data tersebut dihitung dengan rumus metode naïve Bayes.
- Contoh perhitungan data negatif dan data positif seperti pada tabel nomor 2 dan 3.
- Untuk keterangan mengenai tabel 2 dan tabel 3 bisa dilihat pada tabel nomor 4.

Tabel 2. Contoh Perhitungan Data Negatif

untuk negatif	
P(-)	(2/5)
pemilu (-)	$(2+1)/(17+30)=0.0634$
dapat (-)	$(0+1)/(17+30)=0.021$
berjalan (-)	$(0+1)/(17+30)=0.021$
damai (-)	$(1+1)/(17+30)=0.021$
P(-).P(S -)	$(2/5) \times 0.0634 \times 0.021 \times 0.021 \times 0.021$
	<b><math>= 2.349 \times 10^{-7}</math></b>

Tabel 3. Contoh Perhitungan Data Negatif

untuk positif	
P(+)	(3/5)
pemilu (+)	$(2+1)/(13+30)=0.07$
dapat(+)	$(0+1)/(13+30)=0.023$
berjalan (+)	$(0+1)/(13+30)=0.023$
damai (+)	$(1+1)/(13+30)=0.05$
P(+).P(S +)	$(3/5) \times 0.07 \times 0.023 \times 0.023 \times 0.05$
	<b><math>= 1.11 \times 10^{-06}</math></b>

Tabel 4. Keterangan mengenai tabel 2 dan 3

keterangan	
2	jumlah kata pemilu dalam label negatif pada tabel
1	default dari rumus
17	jumlah seluruh kata dalam label negatif
30	jumlah seluruh kata dalam label negatif + label positif

- Setelah didapatkan hasil P x P(S|-) dan P x P(S|+)
- $P \times P(S|-) = 2.349 \times 10^{-7}$
- $P \times P(S|+) = 1.11 \times 10^{-06}$
- Dibandingkan mana yang lebih besar dari kedua hasil tersebut,
- Yang lebih besar adalah yang berlabel positif, maka
- Kalimat dalam testing "pemilu dapat berjalan damai" berlabel **positif**.

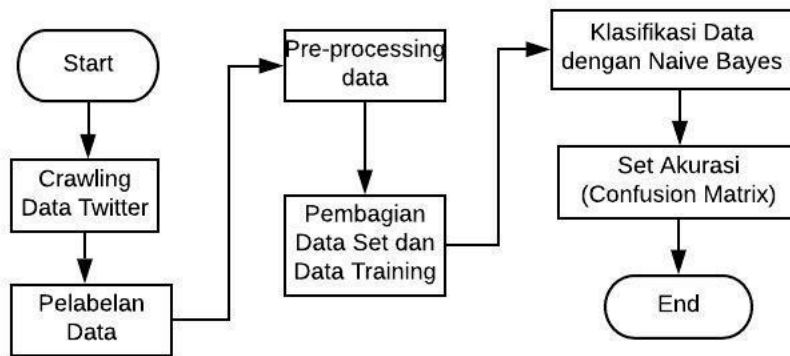
2.4 Analisis Sentimen

Sentimen analisis adalah penambangan kontekstual teks yang mengidentifikasi dan mengekstrak informasi subjektif dalam sumber. Informasi tersebut berguna untuk orang-orang seperti halnya para pebisnis untuk memahami sentimen sosial dari merek, dan produk mereka.

Sebagai contoh dalam tugas akhir ini, sentimen analisis digunakan untuk melihat bagaimana sentimen masyarakat Indonesia khususnya pada media sosial twitter dalam menanggapi pemilihan umum presiden Indonesia 2019.

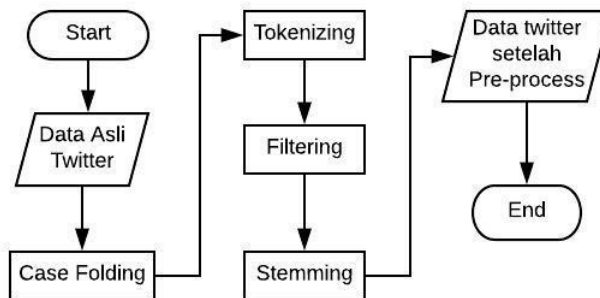
Analisis sentimen mempunyai tiga nilai yaitu, positif, negatif, dan netral [3]. Analisis sentimen positif contohnya, “pemilu Indonesia 2019 pasti aman, jujur, dan adil”. Analisis sentimen negatif contohnya, “saya tidak percaya pada calon nomor 4, memang apa andilnya dalam Negara Indonesia?”. Pada tugas akhir ini, nilai analisis sentimen yang digunakan hanya dua yaitu, positif dan negatif. Analisis sentimen pada data yang didapatkan memperoleh 61% sentimen positif dan 39% mengarah pada sentimen negatif.

3. Perancangan Sistem  
 3.1 Pembangunan model



Gambar 1. Alur Pembangunan Model

Proses pembangunan model pada tugas akhir ini dapat dilihat pada gambar 1 di atas. Proses pertama yang akan dilakukan dalam pengerjaan pembangunan model ini adalah crawling data twitter atau pengumpulan data dari media sosial twitter, lalu seluruh data yang telah didapatkan pada crawling data tersebut dilakukan pelabelan data secara manual sesuai dengan kelas sentimennya yaitu positif atau negatif. Data yang diperoleh dari crawling data twitter dan setelah dilakukan pelabelan berjumlah 500 data. Setelah pelabelan dilakukan proses berikutnya adalah pre-processing. Pre-processing dilakukan untuk membuat data menjadi lebih baik sebelum nanti akan dilakukan klasifikasi[5]. Karena data yang diperoleh dari proses crawling data twitter pasti mempunyai *noise* atau bentuk dari data tersebut tidak terstruktur secara rapi.



Gambar 2. Alur Proses Pada Tahap Pre-processing

Alur mengenai proses *pre-processing* dapat dilihat pada flowchart gambar 2. *Pre-processing* merupakan langkah-langkah yang bertujuan untuk menyiapkan data sesuai dengan kebutuhan secara efisien untuk dapat diproses

dengan suatu metode pada bagian utama sistem [5]. Adapun tahapan pre-processing yaitu: case folding, tokenizing, filtering dan stemming. Adapun penjelasannya sebagai berikut:

3.1.1 Case Folding

Case folding adalah tahapan mengubah semua huruf dalam semua dokumen menjadi huruf kecil.

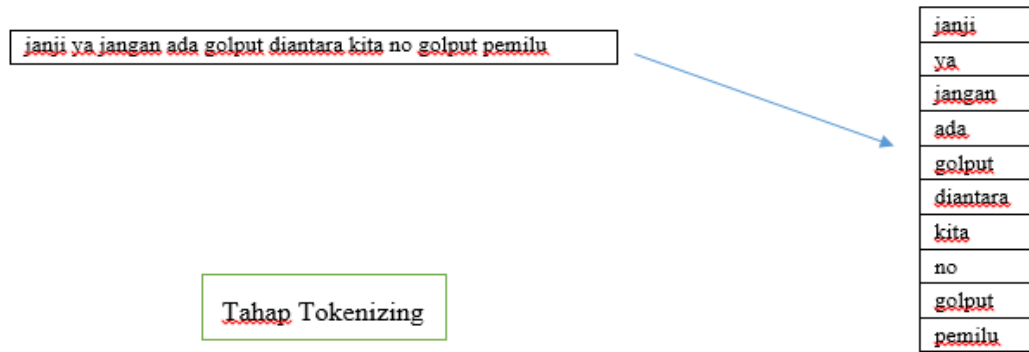
Contoh : - “Janji Ya janGan aDa gOlput Diantara Kita No Golput Pemilu 2019”

Menjadi

- “janji ya jangan ada golput diantara kita no golput pemilu 2019”

3.1.2 Tokenizing

Tokenisasi adalah tahap kedua dalam tahap pre-processing sentimen. Pada tahap ini teks akan dipecah menjadi token-token yang berurutan. Contoh tahap tokenizing dapat dilihat pada gambar nomor 3 berikut :



Gambar 3. Tahap Tokenizing

3.1.3 Filtering

*Filtering* adalah mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma stoplist (membuang kata yang urang penting) atau wordlist (menyimpan kata penting). Stoplist adalah kata-kata yang tidak deskriptif yang dapat dibuang. Contoh kata-kata yang akan dibuang yaitu: “yang, akan, adalah” dan lain-lain.

Contoh tahap filtering dapat dilihat pada gambar nomor 4 berikut :



Gambar 4. Tahap Tokenizing

3.1.4 Stemming

Stemming merupakan suatu proses yang terdapat dalam sistem IR(*Information Retrieval*) yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya dengan menggunakan aturan-aturan tertentu.

Contoh tahap stemming dapat dilihat pada gambar nomor 5 berikut :



Gambar 5. Tahap Tokenizing

3.2 Pengumpulan Data

*Crawling* data twitter adalah pengambilan data dari media sosial twitter. Data yang diambil berupa Contoh hasil dari *Crawling* data twitter dapat dilihat seperti pada gambar nomor 6:

A	B	C
NO	Created-At	Text
1	3/18/19	Setelah Debat Cawapres, mana yang penampilannya paling oke? Pernyataan apa yang paling menarik, tulis di komentar#DebatPilpres20
2	3/18/19	Setelah menyaksikan #DebatCawapres2019, siapa kandidat cawapres yang mempunyai program paling nyata dan Anda sukai? Ayo repl
3	3/19/19	BPN Duga Perangkat Desa Dikumpulkan Karena Jokowi Panik #DebatPilpres2019 #DebatKetigaPilpres2019 <a href="https://t.co/vB4KdBnCqy">https://t.co/vB4KdBnCqy</a>
4	3/26/19	RT @rakyatrukun_ : Sobat rukun. Jangan lupa saksikan debat ketiga Cawapres malam ini pukul 20:00 WIB yang akan membahas tema k
5	3/25/19	RT @kompascom: Kata @sandiuno,Di Jakarta, program OK OCE sudah bisa menurunkan pengangguran sebanyak 20.000 di tahun 2018
6	3/25/19	RT @01UntukRiau: Maju karena Manusia#debatcapres #pilpres2019 #jokowimembangun #jokowiamin #01untukriau #riamemilihjokov
7	3/25/19	RT @01UntukRiau: Pak Kyai emang mantul!#debatcapres #pilpres2019 #jokowimembangun #jokowiamin #01untukriau #riamemilihjol
8	3/25/19	RT @01UntukRiau: Mantul pak Kyaiiii!!!#debatcapres #pilpres2019 #jokowimembangun #jokowiamin #01untukriau #riamemilihjokow
9	3/25/19	RT @01UntukRiau: Telak!#DebatCawapres2019 #DebatPilpres2019 #KyaikuKeren #01untukriau <a href="https://t.co/upjMNYgXFR">https://t.co/upjMNYgXFR</a>
10	3/25/19	RT @jek___ : setuju sama statment pakde "gaji pns/asn sudah cukup, apalagi sudah ada tunjangan kinerja. saatnya merampingkan birokr
11	3/25/19	RT @Kiel_tea: Saya percaya apa yang dilakukan Jokowi-Amin selalu untuk tunas-tunas muda bangsa ini, karena visi-misinya sesuai deng
12	3/25/19	RT @Kiel_tea: Membaca sejarah dan budaya Indonesia di Opera, mantap biar gak perlu lagi lah ke bioskop hehe #Wapreskukyai #Debat
13	3/25/19	RT @stywnyudha: udalah kita semua tuh perna korupsi. manusiawi. yang nga manusiawi itu udah lu bikin baper setelahnya dia malah
14	3/25/19	RT @MetroTVNewsRoom: Berikut "istilah-istilah luar biasa" dari KH. Ma'ruf Amin ketika #DebatCawapres2019 lalu.Tonton terus #Metro

Gambar 6. Contoh Data Awal Yang Diambil Dari Twitter

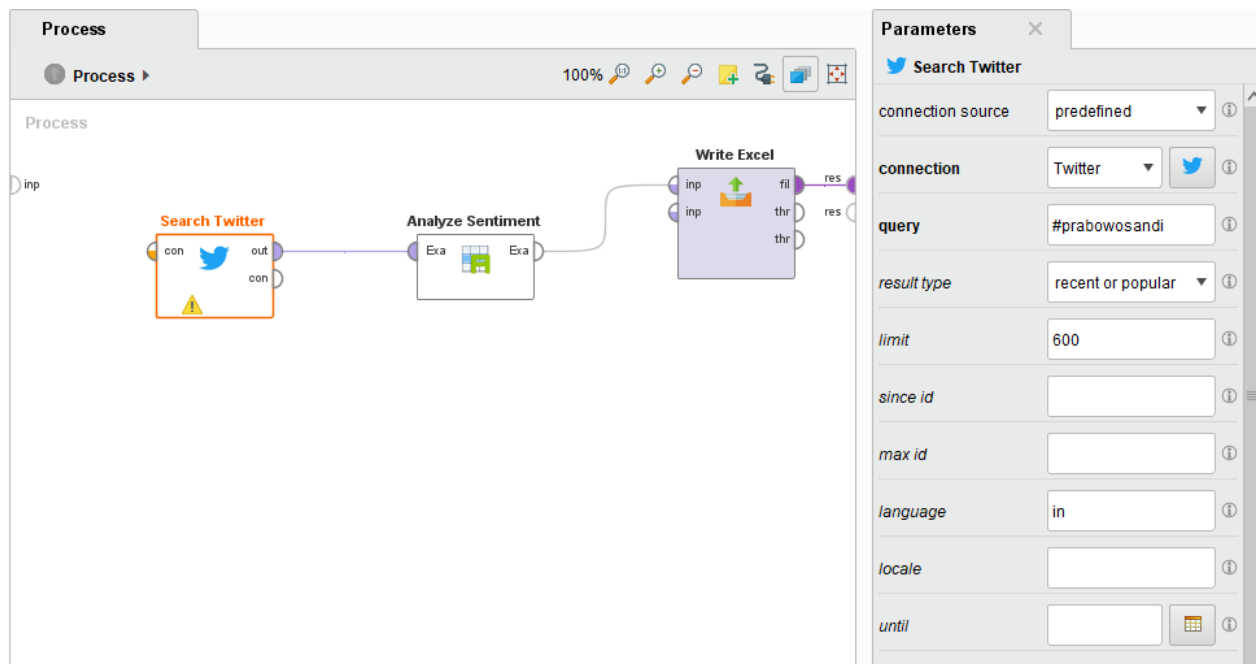
3.2.1 Langkah-langkah crawling data twitter menggunakan rapid miner 9.0 :

Langkah-langkah pengerjaan crawling data twitter menggunakan rapid miner 9.0 :

1. Buka aplikasi rapid miner 9.0
2. Klik menu extensions dan pilih marketplace (Update and Extensions)
3. Pada menu Search, ketik "text", cari text analysis by AYLIEN 0.2.0, centang untuk Select for installation, lalu install 1 packages
4. Pilih Text Analysis by AYLIEN 0.2.0 centang "I accept the terms" lalu install
5. Setelah terinstall maka akan dilakukan restart untuk rapid miner 9.0
6. Setelah masuk kembali, pilih menu operators ketik "twit", dan pilih search twitter tahan dan geser pada layar proses. Pada menu parameters (kanan layar proses) klik ikon "twitter".

7. Add connection, isi name dengan twitter dan create
8. Pilih request access token pada kolom access token, setelah itu akan diteruskan pada link extension dari rapidminer. Isikan username dan password lalu klik authorize app.
9. Setelah itu akan muncul pin, copy lalu paste pada copy code, dan klik “complete”. Klik test dan save all changes.
10. Pada fitur parameters, pilih twitter pada kolom connection, ketik apa yang ingin dicari pada kolom query, ketik jumlah tweet yang ingin didapatkan pada kolom limit, dan bahasa pada language (jika ingin menggunakan bahasa Indonesia maka ketikkan “in”).
11. Pada fitur operators ketik *sentiment*, akan muncul jendela manage connection lalu create connection pada add, ganti name menjadi *sentiment*, setelah itu create.
12. Kita membutuhkan app id dan app key, maka dari itu pada tombol bantuan app id klik f3 lalu klik “here”, nanti akan muncul link pada browser, ikuti petunjuk pengisian “Text Analysis API”, setelah semua diisi dan sign up. Buka email untuk memverifikasinya, nanti app id, dan app key akan muncul. Isi id dan key pada kolom yang tertera di rapidminer sebelumnya, dan simpan dengan klik *save all changes* [4].
13. Hubungkan panel “search twitter” dengan “*Analyze Sentiment*”, serta “*Analyze Sentiment*” dengan res.
14. Jalankan program dengan klik tombol play, tunggu hingga proses selesai.
15. Setelah selesai, maka data pengambilan data dari twitter tersebut akan muncul pada layar.

Berikut merupakan tampilan proses crawling data twitter pada rapid miner, dapat dilihat pada gambar nomor 7.



Gambar 7. Proses crawling data twitter pada rapid miner



Kata kunci untuk crawler (pengumpulan data) seperti pada tabel 5 berikut :

Tabel 5. Kumpulan kata kunci

Kumpulan kata kunci
#Pilpres2019
#DebatPilpres2019
#pemilu2019
#jokowi
#prabowo
jokowi pemilu
prabowo pemilu

3.2.2 Kuisisioner

TEST	KUISISIONER
	CONTOH. Berikut pernyataannya
	Politikus kok baper? #DebatCapres #DebatPilpres2019.....
	Anda bisa memberikan nilai pernyataan berikut dengan 1 untuk positif, jika anda rasa pernyataan diatas mengandung kalimat positif (tidak kontra)
	Tolong kasih nilai terhadap pernyataan diatas pada kolom "NILAI" di sebelah kiri "pernyataan".
	1 untuk positif
	2 untuk negatif
NILAI	PERNYATAAN
1	1 Janji yaa!! Jangan ada golput diantara kita #NoGolput #Pemilu2019 #SayNoToGolput
2	1 Presiden yg baik bisa memberikan pengaruh yg baik, yg malas menjadi rajin, yg korupsi menjadi tidak korupsi, yg saling benci menjadi saling cinta, dan itu tidak terliha
3	2 Banyak jg warganegara yg mau golput...ah ayolah nyoblos besok 17 April. Seenggaknya kl bego jgn dikeluarin bgt, soalnya org bego tubuh golput. seenggaknya nya kl
4	2 Wahai cebong dan kampret, bersatulah! Berubah jadi 'cepret'
5	2 pengen masa pilpres cepet selesai, bosen liat dikit-dikit ngatain cebong kampret cebong kampret bgst klen
6	2 Kalo yang ngomong cebong kampret gw sanksi. Tapi kalo yang ngomong faisal basri gw percaya
7	1 Pemilu & Pilpres adalah pesta demokrasi. Layaknya sebuah pesta, semua bergembira, bukan marah2, maki2 dan gebrak2 meja. Pesta macam apa itu?
8	1 Sebel nama grup nya " alumni SMP " bahas nya pilpres mulu. Nyinyir, Ga siang ga malem, gua out aja ga guna tuh grup tujuannya apa kali
9	2 jelang pemilu semakin banyak politikus2 tolol yg meretweet akun2 gk jelas penyebar hoax.. dah gt mengatasnamakan ra...
10	2 Eh cebong emg lo tau apa yg prabowo lakukan 5thn blakangan? Beliau mah ikhlas g perlu pencitraan depan kamera apa yang dilakukan.tapi ini karena pilpres kudu pada
11	1 Makin dekat voting day sering nyinyir ttg amplop. Apa ga dibagi sama temen lo yg sering bahas usai debat pilpres?
12	2 Ini mulut apa selet kerbau semoga pasca pilpres ada tindakan tegas terhadap monyet satu ini

Gambar 8. Kuisisioner

3.2.2.1 Tahap Kuisisioner

Pada pengerjaan tugas akhir ini mempunyai tahap kuisisioner. Tahap ini dilakukan ketika data dari twitter sudah didapatkan dan dilakukan proses *pre-processing*. Tahap kuisisioner ini atau bisa disebut dengan tahap survey adalah tahap untuk mengetahui bagaimana masyarakat sekitar menilai tentang topik tugas akhir ini yang berjudul Analisis Sentimen mengenai Pemilu Indonesia 2019. Contoh data kuisisioner pada tugas akhir ini dapat dilihat pada gambar 8 di atas. Data dari twitter mengenai pemilu Indonesia 2019 yang berjumlah 600 record dibagikan kepada 10 orang. Setelah dibagikan, para responden tersebut memberi nilai pada setiap record dengan nilai positif dengan angka 1 dan negative dengan angka 2.

Setelah didapatkan semua jawaban dari responden, data tersebut diolah untuk mengetahui nilai dari setiap record. Data dengan label positif dan negative inilah yang disebut dengan data aktual. Data aktual tersebut akan digunakan untuk pengerjaan data menggunakan metode naïve Bayes. Data aktual tersebut juga digunakan untuk membuat tabel *confusion matrix*.

3.3 Skenario Pengujian Naïve Bayes Classifier

Data set yang berjumlah 500 dibagi ke dalam dua kategori yaitu data training dan data set. Dan sebelum pengujian dilakukan maka terlebih dahulu dilakukan skenario pengujian seperti yang tertera pada gambar 9.0. Skenario pengujian dilakukan untuk melihat perbandingan jumlah data yang akan dimasukkan pada dua kategori yaitu data training dan data testing [11]. Berikut adalah tabel untuk skenario pengujian dalam bentuk persentase.

Tabel 6. Jumlah data berdasarkan label positif dan negatif

No.	Label	Jumlah Tweet
1	Positif	365
2	Negatif	235

Jumlah porsi data training dan data testing untuk skenario pengujian dapat dilihat pada tabel 7 berikut :

Tabel 7. Jumlah porsi data training dan data testing

No.	Data Training	Data Testing
1	55 % (330 data training)	45 % (270 data testing)
2	60 % (360 data training)	40 % (240 data testing)
3	70 % (420 data training)	30 % (180 data testing)
4	80 % (480 data training)	20 % (120 data testing)
5	90 % (540 data training)	10 % ( 60 data testing)

### 3.5 Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining atau Sistem Pendukung Keputusan [12]. Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Nilai *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive* (TP) merupakan data positif yang terdeteksi benar. *False Negative* (FN) merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif. Presisi adalah data yang diambil berdasarkan informasi yang kurang. Dalam klasifikasi biner, presisi dapat dibuat sama dengan nilai prediksi positif.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{2.1}$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\% \tag{2.2}$$

$$\text{Recall} = \frac{TP}{FN+TP} \times 100\% \tag{2.3}$$

- TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
- FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem

## 4. Hasil dan Analisis

### 4.1 Skenario Pengujian

#### 4.1.1 Skenario Pengujian 1 (55% : 45 %)

Perbandingan data training dan data testing : 330 data training dan 270 data testing.

Untuk jumlah data skenario pengujian 1 bisa dilihat pada tabel 8 berikut :

Tabel 8. Skenario pengujian 1

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	103	11
Negatif	28	38

## 4.1.2 Skenario Pengujian 2 (60% : 40 %)

Perbandingan data training dan data testing : 360 data training dan 240 data testing.  
Untuk jumlah data skenario pengujian 2 bisa dilihat pada tabel 9 berikut :

Tabel 9. Skenario pengujian 2

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	71	6
Negatif	16	27

## 4.1.3 Skenario Pengujian 3 (70% : 30%)

Perbandingan data training dan data testing : 420 data training dan 180 data testing.  
Untuk jumlah data skenario pengujian 3 bisa dilihat pada tabel 10 berikut :

Tabel 10. Skenario pengujian 3

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	124	65
Negatif	49	32

## 4.1.4 Skenario Pengujian 4 (80% : 20%)

Perbandingan data training dan data testing : 480 data training dan 120 data testing.  
Untuk jumlah data skenario pengujian 4 bisa dilihat pada tabel 11 berikut :

Tabel 11. Skenario pengujian 4

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	142	27
Negatif	42	29

## 4.1.5 Skenario Pengujian 5 (90% : 10%)

Perbandingan data training dan data testing : 540 data training dan 60 data testing.  
Untuk jumlah data skenario pengujian 5 bisa dilihat pada tabel 12 berikut :

Tabel 12. Skenario pengujian 5

Kelas Aktual	Kelas klasifikasi	
	Positif	Negatif
Positif	37	6
Negatif	6	11

4.1.1.1 Hasil Confusion Matrix

Hasil perhitungan confusion matrix bisa dilihat pada tabel 13 berikut :

Tabel 13. Tabel Hasil Confusion Matrix

Confusion Matrix	Porsi Data Training dan Data Testing				
	55 % : 45% (330 : 270)	60 % : 10 % (360 : 240)	70 % : 30% (420 : 180)	80 % : 20% (480 : 120)	90 % : 10 % (540 : 60)
Akurasi	63.33	64.17	70	70.42	71.67
Presisi	78.63	81.61	71.68	77.17	86.04
Recall	73.05	72.45	79.49	83.04	77.08

4.3 Analisis

Berdasarkan hasil pengujian dari data di atas, disimpulkan bahwa sistem ini dapat menghasilkan akurasi yang cukup baik dengan tingkat akurasi sebesar 71.67 % menggunakan *confusion matrix*. Pada Proses Pre-processing data detail dari setiap prosesnya yaitu case folding, tokenizing, filtering dan stemming sangat berpengaruh terhadap proses yang dilakukan selanjutnya. Dari proses yang sudah dilakukan pada tahap skenario pengujian didapatkan bahwa semakin banyak data training yang dipakai maka semakin besar tingkat akurasinya. Dapat dilihat pada tabel Confusion Matrix bahwa tingkat akurasi mengalami penurunan ketika porsi data training semakin mengecil.

5. Kesimpulan

Hasil dari pengamatan sentimen masyarakat mengenai pemilu presiden yang dilaksanakan pada tahun 2019 menunjukkan peningkatan jumlah *tweet* pada jejaring sosial Twitter ketika rangkaian proses pemilu dilaksanakan. Terlihat mulai menunjukkan antusias mengenai pemilu, seperti contohnya ketika rangkaian debat dimulai. Dari percakapan *tweet* tersebut, bisa diambil informasi mengenai seberapa antusias masyarakat dalam melihat pemilu presiden, dan disamping itu juga bisa diambil informasi mengenai sentimen positif dan sentimen negatif terhadap rangkaian proses pemilu itu sendiri, terhadap calon presiden dan wakil presiden, serta menunjukkan bagaimana kesediaan masyarakat untuk mengikuti pencoblosan pada tanggal 17 april 2019.

Hasil dari pengamatan polaritas sentimen masyarakat terhadap rangkaian proses pemilu, dan terhadap calon presiden dan wakil presiden adalah sebesar 61% menunjukkan sentimen positif dan 39% menunjukkan sentimen negatif. Melalui perhitungan metode naïve Bayes, dan dilanjutkan dengan metode *confusion matrix* menunjukkan bahwa 71.67% perhitungan akurasi didapat. Hal ini menunjukkan bahwa perhitungan proses sentimen analisis menggunakan metode naïve Bayes menghasilkan hasil yang cukup baik.

## Daftar Pustaka

- [1] Sembodo, J. E., Setiawan, E. B., & Baizal, Z. K. A. (2017). A Framework For Classifying Indonesian News Curator In Twitter. *TELKOMNIKA (Telecommunication Computing Electronics And Control)*, 15(1), 357-364.
- [2] Nurhuda, Faishol. (2013). Analisis Sentimen Masyarakat Terhadap Calon Presiden Indonesia 2014 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal ITSMART*.
- [3] Liu, B. (2012). *Sentiment Analysis And Opinion Mining*. Morgan & Claypool Publishers.
- [4] S, B. F. (2014). *[Pengenalan] Apa Itu Twitter API Dan Pembuatan Consumer Key Dan Consumer Secret ?* Retrieved Januari 2016, 21, From
- [5] Huracans. (2017). Tahap Preprocessing Text Mining. Huracans Blog 27 September 2017. [Http://Frosthuracan.Blogspot.Com/2017/09/Tahap-Preprocessing-Text-Mining.Html](http://Frosthuracan.Blogspot.Com/2017/09/Tahap-Preprocessing-Text-Mining.Html)
- [6] Informatikalogi (2017). Algoritma Naïve Bayes. Informatikalogi 15 Juli 2017. [Https://Informatikalogi.Com/Algoritma-Naive-Bayes/](https://Informatikalogi.Com/Algoritma-Naive-Bayes/)
- [7] Dwata, Apriandi. (2016). Analisis Sentimen Pelanggan Wifi.Id Pada Twitter Dengan Support Vector Machine. (*Jurnal Sistem Informasi Dan Bisnis Cerdas*). 1 Februari 2016.
- [8] Jurafsky, Daniel, Dan Martin, James. (2017). Naïve Bayes And Sentimen Classification. Naïve Bayes Classifier. 7 Agustus 2017. (Chapter 6)
- [9] Ling, Juen, I Putu Eka, Dan Tjokorda Bagus. Analisis Sentimen Menggunakan Metode *Naïve Bayes Classifier* Dengan Seleksi Fitur *Chi Square*. *Jurnal Matematika*. 3 Agustus 2014.
- [10] Pamungkas, Diyarsa Singgih. (2015). Analisis Sentiment Pada Sosial Media Twitter Menggunakan Naive Bayes Classifier Terhadap Kata Kunci “Kurikulum 2013”. *Techno.COM*. 4 November 2015.
- [11] Pratama, Novan Dimas. (2018). Analisis Sentimen Pada *Review* Konsumen Menggunakan Metode *Naive Bayes* Dengan Seleksi Fitur *Chi Square* Untuk Rekomendasi Lokasi Makanan Tradisional. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*. 9 September 2018.
- [12] Teori Kuliah Komputer. (2018). Pengujian Dengan Confusion Matrix. 23 Juli 2018. <http://www.kuliahkomputer.com/2018/07/pengujian-dengan-confusion-matrix.html>