

**SENTIMEN ANALISIS DATA TWITTER TERHADAP  
CALON WAKIL PRESIDEN 2019 SANDIAGA SALAHUDDIN UNO**

**Yoga Religia, Heri Purwanto**

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa  
yogareligia@pelitabangsa.ac.id

Disetujui, 25 Maret 2019

**Abstraksi**

Media social *twitter* merupakan salah satu contoh media social yang digunakan masyarakat untuk saling berinteraksi satu sama lain. *Twitter* memberikan layanan kepada penggunanya untuk mengirim dan membaca *tweets* yang telah dibagikan, sehingga masyarakat lebih memilih menuangkan opininya melalui media social dari pada menyampaikannya secara langsung. Opini masyarakat yang tertuang dalam media social *twitter* berupa sebuah persepsi, baik itu positif maupun negative. Melimpahnya opini masyarakat dapat dimanfaatkan sebagai bahan penelitian untuk mencari sebuah informasi. Pemanfaatan informasi tersebut membutuhkan teknik analisis yang tepat sehingga informasi yang dihasilkan mampu membantu banyak pihak dalam mengambil keputusan. Penggunaan teknik dalam pengolahan data dapat diselesaikan teknik analisa sentiment atau *opinion mining*. Oleh sebab itu, pada penelitian ini mencoba melakukan analisa sentiment untuk melihat persepsi masyarakat terhadap calon wakil presiden 2019 sandiaga salahudin unno dari partai gerindra pada media social *twitter* menggunakan metode Naïve Bayes Classifier dengan mengklasifikasikan sentimen menjadi positif, negative dan netral.

**Kata Kunci:** Analisa Sentiment, Naïve Bayes Classifier, Persepsi, Twitter

**Abstract**

*Twitter social media is one example of social media that is used by the community to help one another. Twitter provides a service for its users to send and read the tweet that has been shared, so that the public prefers to express their opinions through social media than when delivering it directly. Citizens stated in Twitter social media gave their opinions, both positive and negative. Can be used as research material to search for information. Utilization of this information requires appropriate analysis techniques required information generated requires the help of parties in retrieving results. The use of techniques in data processing can be completed sentiment analysis techniques or opinion mining. Therefore, this research tries to do sentiment analysis to see the public perception of the 2019 vice presidential candidate sandiaga Salahudin unno from the Gerindra party on Twitter social media using the Naïve Bayes Classifier method by classifying sentiments looking for positive, negative and neutral.*

**Keywords:** Sentiment Analysis, Naïve Bayes Classifier, Perception, Twitter

## Pendahuluan

Perkembangan akses teknologi informasi yang semakin pesat pada era globalisasi saat ini dapat memberikan kemudahan untuk berkomunikasi lebih efektif dan efisien, contohnya adalah internet. Seperti halnya yang telah diungkapkan oleh Kementerian Komunikasi dan Informatika (kemenkominfo, 2017) bahwa pengguna internet di Indonesia saat ini mencapai 63 juta orang. Dari angka tersebut, terdapat 95 persennya menggunakan internet untuk mengakses media sosial. Salah satu media sosial yang diakses masyarakat adalah Twitter.

Twitter adalah situs microblogging populer dimana pengguna membuat status yang disebut "tweet". Tweet memiliki batas maksimal 140 karakter. Orang memposting pesan singkat, menggunakan berbagai bentuk singkatan, menggunakan emoticon dan karakter lain yang mengekspresikannya arti khusus dari kalimat tersebut (Kotwal et al., 2016). Tweet atau pesan yang dibagikan di twitter biasanya merupakan topik yang sedang hangat dibicarakan dan kadang menjadi trending topic di twitter.

Seperti akhir – akhir ini yang menjadi topic berita yaitu isu mengenai calon wakil presiden dari partai gerindra untuk pemilihan presiden 2019 mendatang yaitu Sandiaga Salahuddin Uno. Memang semua yang berhubungan dengan pemilihan presiden menjadi banyak di perbincangkan di media sosial. Sehingga bermunculan pendapat positif maupun negatif. dari masyarakat terhadap isu tersebut. Pendapat-pendapat tersebut bisa diolah dan dianalisa agar menghasilkan data atau informasi yang sehingga informasi yang dihasilkan dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan.

Salah satu teknik pemrosesan teks yang tepat yaitu analisa sentimen. Analisis sentimen adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik (Medhat & Korashy 2014).

Dalam penelitian ini, peneliti mencoba melakukan analisa sentimen untuk melihat persepsi masyarakat terhadap calon wakil presiden 2019 partai gerindra Sandiaga Salahuddin Uno pada media sosial twitter menggunakan metode naive bayes, dengan mengklasifikasikan sentimen menjadi positif, negatif dan netral. Selain itu juga untuk melihat keakurasian metode yang digunakan sebelum dan setelah diterapkan fitur ekstraksi n-gram karakter. Dan data dari hasil penelitian ini bisa dimanfaatkan untuk lembaga terkait sebagai bahan pertimbangan kembali dalam mengambil keputusan atau membuat kebijakan yang menyangkut hajat orang banyak.

## 2. Tinjauan Studi

### 2.1. Text minning

Text Mining adalah proses otomatis untuk mendeteksi dan mengungkapkan pengetahuan baru dan tidak terungkap serta hubungan dan pola dalam sumber data teks yang tidak terstruktur (younis :2015).

Menurut Feldman dan Sanger (2007:10) “Text Mining adalah area penelitian baru dan menari yang mencoba memecahkan masalah informasi dalam jumlah besar (overload information) dengan menggunakan teknik data mining, machine learning, natural language processing (NLP), information retrieval (IR), dan knowledge management. Text mining melibatkan preprocessing dokumen (text categorization, information extraction, term extraction)”. Sedangkan menurut Kumar dan Bhatia (2013) “Text mining adalah bidang baru yang mencoba untuk menggali informasi yang berguna dari teks bahasa alami (natural language). Hal ini dapat diartikan sebagai proses menganalisa teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu”.

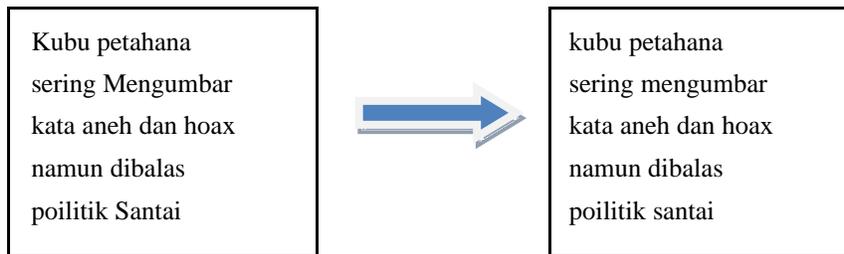
Berdasarkan pendapat para ahli tersebut di atas bisa disimpulkan bahwa pada dasarnya proses kerja dari text mining sama dengan data mining yaitu menggali informasi dari sekumpulan data dalam jumlah besar untuk tujuan tertentu, namun yang menjadi perbedaan adalah pola yang digunakan oleh text mining diambil dari sekumpulan data yang tidak terstruktur atau semi terstruktur data sedangkan dalam data mining pola yang diambil dari database data terstruktur. Tujuannya intinya dari text mining adalah mengubah teks (data tidak terstruktur) menjadi data format terstruktur untuk analisis, melalui serangkaian proses pengolahan teks.

Di dalam text mining sebelum data tekstual yang tidak terstruktur melalui proses text mining, kumpulan data tekstual tersebut harus melalui proses text preprocessing. Text Preprocessing ini bertujuan merubah data yang tidak terstruktur menjadi semi terstruktur ataupun terstruktur sebelum dilakukan proses text mining berikutnya

### 2.2. Text preprocessing

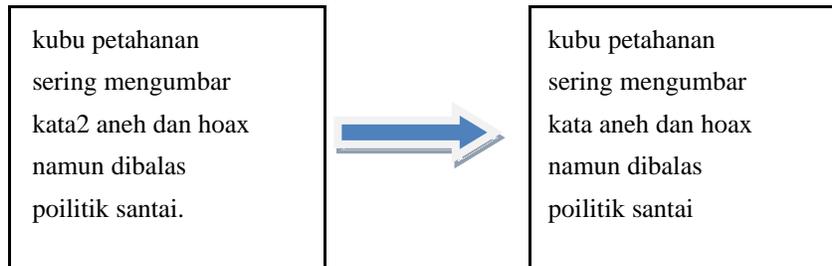
Dalam text preprocessing terdiri dari beberapa tahapan sebagai berikut:

- 1) *Cleansing* adalah suatu tahap di mana karakter maupun tanda baca yang tidak diperlukan dibuang dari teks. Contoh karakter yang dibuang adalah tanda seru, tanda tanya, koma dan titik (Hadna et al. , 2016).
- 2) *Case folding*, yaitu penyeragaman bentuk huruf serta penghapusan angka dan tanda baca. Dalam hal ini yang digunakan hanya huruf latin antara a sampai dengan z (Nur dan Santika 2011). Contoh dari tahap *case folding* sebagai berikut:



Gambar 1. Contoh Proses *Case Folding*

3) *Tokenizing*, pada proses tokenize ini, semua tanda baca, simbol, atau apapun yang bukan huruf dihilangkan sehingga menjadi sekumpulan kata secara utuh (Indrayuni, 2016). Contoh dari *tokenizing* sebagai berikut :



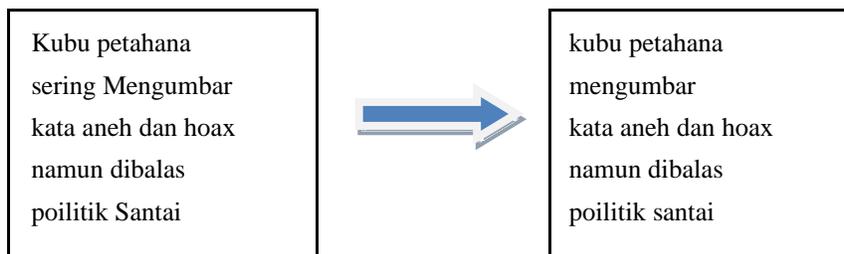
Gambar 2. Contoh Proses *Tokenizing*

4) *Convert Word*, pentingnya *convert word* adalah untuk mengkonversi kalimat yang tidak baku, saat ini penggunaan kalimat alay atau bahasa gaul mengakibatkan penggunaan bahasa indonesia tidak baku (Mujilahwati, 2016).

### 2.3. Seleksi Fitur (*Feature Selection*)

#### a) *Stopword Removal*

*Stopwords* dapat diartikan sebagai menghilangkan kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan (Ling et al. 2014). Contoh stopwords dalam bahasa indonesia seperti “yang”, “juga”, “dari”, “dia”, “kami”, “kamu”, “aku”, “saya”, “ini”, “itu”, “atau” dan lain sebagainya. Pada proses ini menghapus kata-kata yang sama dengan daftar kata-kata ditabel stopwords. Berikut contoh proses *stopword removal*.



Gambar 3. Contoh Proses *Stopword Removal*

#### b) *Stemming*

*Stemming* menurut (Ling et al, 2014) yaitu “*Stemming* merupakan salah satu proses dari mengubah token yang berimbuhan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada token tersebut”. Pentingnya *stemming* dalam proses pembuatan sistem adalah untuk menghilangkan imbuhan pada awalan dan akhiran. *Stemming* bertujuan mengurangi variasi kata yang memiliki kata dasar sama (Nur & Santika, 2011). Sebagai contoh, kata menyenangkan, menyenangkan dapat dikatakan serupa atau satu kelompok da dapat diwakili oleh satu kata umum senang.

Algoritma *Stemming* adalah prosedur komputasi yang mencari asal kata dari suatu kata dalam kalimat yang dilakukan dengan cara memisahkan masing-masing kata dari kata dasar dan imbuhan. Pada saat ini ada beberapa algoritma *stemming* untuk Bahasa Indonesia yang telah dikembangkan, salah satunya yaitu algoritma Nazief dan Andriani.

**2.4. N-Gram (Feature Extraction)**

*N-gram* adalah potongan n karakter dalam suatu string tertentu atau potongan n kata dalam suatu kalimat tertentu. Misalnya dalam kata “Belajar” akan didapatkan *n-gram* sebagai berikut:

**Tabel 1. Contoh Pemotongan n-gram Berbasis Karakter**

N-gram	N-gram Karakter
Unigram	B,E,L,A,J,A,R
Bigram	_B, BE, EL, LA, AJ, JA, R_
Trigram	_BE, BEL, ELA, LAJ, AJA, JAR, AR_, R_ _

Contoh pemotongan N-gram berbasis karakter : Karakter blank “\_” digunakan untuk merepresentasikan spasi di depan dan di akhir kata (Sadida et al. 2017). Dapat disimpulkan bahwa untuk string berukuran n akan dimiliki n unigram dan n+1 bigram, n+1 trigram, n+1 quadgram dan seterusnya. Penggunaan N-gram untuk matching kata memiliki keuntungan sehingga dapat diterapkan pada recovery pada input karakter ASCII yang terkena noise, interpretasi kode pos, information retrieval dan berbagai aplikasi dalam pemrosesan bahasa alami.

Keuntungan N-gram dalam matching string adalah berdasarkan karakteristik n-gram sebagai bagian dari suatu string, sehingga kesalahan pada sebagian string hanya akan berakibat perbedaan pada sebagian n-gram. Sebagai contoh jika n-gram dari dua string dibandingkan, kemudian kita menghitung cacah n-gram yang sama dari dua string tersebut maka didapatkan nilai similaritas atau kemiripan dua string tersebut yang bersifat resistan terhadap kesalahan tekstual.

Kemiripan antara kata JOKO dengan JOKI (ada perbedaan 1 huruf), dapat diukur derajat kesamaan dengan cara menghitung berapa buah n-gram yang diambil dari dua kata tersebut yang bernilai sama, yaitu:

JOKO : \_J, JO, OK, KO, O\_  
 JOKI : \_J, JO, OK, KI, I\_  
 kesamaan : 3.

Sementara antara kata JOKO dengan JONI (ada perbedaan 2 huruf), nilai kesamaan adalah :

JOKO : \_J, JO, OK, KO, O\_  
 JONI : \_J, JO, ON, NI, I\_  
 kesamaan : 2.

Sehingga dapat disimpulkan bahwa kemiripan atau kesamaan antara JOKO-JOKI daripada antara JOKO-JONI (Hamzah, 2010).

**2.5. Analisa Sentiment**

Menurut Medhat dari Korashy (2014) ”Analisa sentimen atau *opinion mining* adalah studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas tersebut dapat menggambarkan individu, kejadian atau topik”.

Jika diberikan suatu set dokumen teks yang berisi opini mengenai suatu objek, maka *opinion mining* bertujuan untuk mengekstrak atribut dan komponen dari objek yang telah dikomentasi pada setiap dokumen dan untuk menentukan apakah opini tersebut bermakna positif atau negatif. *Sentiment Analysis* dapat dibedakan berdasarkan sumber datanya. Berdasarkan level sumber datanya *Sentiment Analysis* terbagi menjadi 2 kelompok besar yaitu (Clayton. 2011) :

1) *Coarse-grained Sentiment Analysis*

*Sentiment Analysis* yang dilakukan adalah pada level dokumen. Secara garis besar fokus utama dari *Sentiment Analysis* jenis ini adalah menganggap seluruh isi dokumen sebagai sebuah sentimen positif atau sentimen negatif.

2) *Fined-grained Sentiment Analysis*

*Sentiment Analysis* yang dilakukan pada level kalimat. Fokus utama *fined-grained Sentiment Analysis* adalah menentukan sentimen pada setiap kalimat.

## 2.6. Twitter

*Twitter* adalah situs *microblogging* populer dimana pengguna membuat status yang disebut “*tweets*”. *Tweets* memiliki batas maksimal 140 karakter. Orang memposting pesan singkat, menggunakan berbagai singkatan, menggunakan *emoticon* dan karakter lain yang mengekspresikannya arti khusus dari kalimat tersebut, (Kotwal et al, 2016). Situs *microblogging* telah berevolusi menjadi sebuah sumber beragam jenis informasi.

Hal ini disebabkan oleh sifat *microblogging* dimana orang menulis pesan *realtime* tentang pendapat mereka mengenai berbagai topik, mendiskusikan isu terkini, mengeluh, dan mengungkapkan sentimen positif untuk produk yang mereka gunakan dalam kehidupan sehari-hari. Data tersebut dapat dimanfaatkan secara efisien untuk pemasaran atau studi sosial.

Dengan menggunakannya, pengguna akan mudah untuk mengikuti *trend*, cerita, informasi dan berita dari seluruh penjuru dunia. Selain itu, *Twitter* juga membantu penggunanya untuk selalu terhubung dengan orang-orang terdekatnya. Ketika penggunanya mengirimkan *tweet*, pesan tersebut bersifat publik dan bisa dikases oleh siapapun, dimanapun, dan kapanpun. Bahkan, bagi orang-orang yang mengikuti (*follow*) akun *Twitter* tersebut, *tweet* tersebut akan secara otomatis muncul di timeline orang tersebut. Berikut ini adalah beberapa istilah yang dikenal dalam *Twitter* (Hadna et al, 2016) :

1) *Mention*

*Mention* adalah menyebut atau memanggil pengguna *Twitter* lain dalam sebuah *tweet*. *Mention* dilakukan dengan menuliskan “@” diikuti dengan nama pengguna lain.

2) *Hashtag*

*Hashtag* digunakan untuk menandai sebuah topik pembicaraan di *Twitter*. Penulisan *hashtag* dimulai dengan tanda “#” diikuti dengan topik yang sedang dibahas. *Hashtag* biasa digunakan untuk meningkatkan visibilitas *tweet* pengguna.

3) *Emoticon*

*Emoticon* adalah ekspresi wajah yang diekspresikan dengan kombinasi antara huruf, tanda baca, dan angka. Pengguna biasa menggunakan *emoticon* untuk mengekspresikan *mood* yang sedang mereka rasakan.

4) *Trending Topics*

Jika *hashtag* adalah cara untuk menandai sebuah topik pembicaraan di *Twitter*, maka *trending topics* adalah kumpulan dari topik pembicaraan yang sangat populer di *Twitter*.

## 2.7. Naive Bayes Classification

Naive Bayes Classifier merupakan teknik pembelajaran algoritma data mining yang memanfaatkan metode probabilitas dan statistik. Naive Bayes Classifier dalam melakukan klasifikasi terdapat dua proses penting yaitu *learning (training)* dan *testing*. Penelitian ini menggunakan *data training* dari opini masyarakat yang diperoleh yaitu sebanyak 350 buah dengan kategori opini positif 150 buah, opini negatif 150 buah dan opini netral 50 buah beserta klasifikasinya secara manual atau data yang sudah diketahui kategorinya. Sedangkan *data testing* digunakan untuk prediksi klasifikasi atau data yang belum diketahui klasifikasinya.

Proses klasifikasi Naive Bayes Classifier terhadap dokumen yaitu dengan mempresentasikan setiap dokumen dengan atribut “ 1, 2, 3 , ... , ” yang mempunyai makna bahwa 1 untuk kata pertama, 2 adalah kata kedua, dan seterusnya. Untuk himpunan kategori *tweet* dipresentasikan dengan . Saat melakukan proses klasifikasi dokumen, Naive Bayes Classifier akan mencari nilai probabilitas tertinggi dari:

## 2.8. Rapid Miner

*RapidMiner* adalah salah satu *software* untuk pengolahan *data mining*. Pekerjaan yang dilakukan oleh *RapidMiner text mining* adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah.

*RapidMiner* menyediakan prosedur data mining dan machine learning, di dalamnya termasuk: ETL (*extraction, transformation, loading*), data *preprocessing*, *visualisasi*, *modelling* dan evaluasi. Proses data *mining* tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI. Penyajiannya dituliskan dalam bahasa pemrograman Java.

### 3. Kerangka Konsep

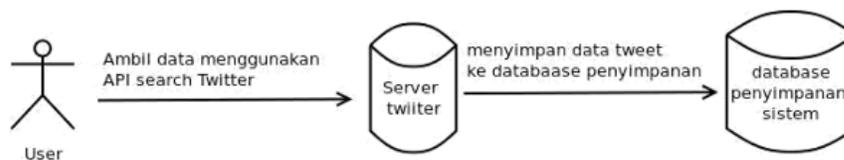
Penerapan native bayes, pada penelitian ini merujuk kepada teori yang digunakan oleh Adi Kusumah (2017). Penelitiannya tentang algoritma naive bayes mampu mencapai tingkat akurasi 89.67%. dan pengaruh ekstraksi fitur *N-Gram* yang diterapkan mampu meningkatkan nilai akurasi pada algoritma *naive bayes* sekitar 2,33% dalam pengklasifikasian data tweet yang di ukur dengan *rapidminer*. Kemudian Penelitian yang dilakukan oleh Fauziah Afshoh (program studi informatika fakultas komunikasi dan informatika Universitas Muhammadiyah Surakarta 2017) dengan judul analisa sentimen menggunakan *naive bayes* untuk melihat persepsi masyarakat terhadap kenaikan harga jual rokok pada media sosial *twitter* berdasarkan topik kenaikan harga jual rokok pada media sosial *twitter* Hasil analisis dan pengujian yang telah dilakukan menggunakan data latih sebanyak 350 buah dari opini masyarakat mengenai topik kenaikan harga jual rokok menggunakan metode *Lexicon Based* menghasilkan nilai *precision* 53% pada data opini negatif dan *recall* sebesar 57% pada data opini negatif Sedangkan untuk evaluasi kinerja sistem menggunakan Naive Bayes.

Selanjutnya, penelitian ini pun merujuk kepada penelitian yang dilakukan oleh Muljono, Dian Putri Artanti, Abdul Syukur Adi Prihandono, De Rosal Moses Setiadi ( Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang 2018) dengan judul Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma *Naive Bayes* Paper ini mempresentasikan analisis sentimen terhadap data opini di twitter di beberapa situs online marketplace di Indonesia. Metode untuk melakukan analisis sentimen menggunakan algoritma pengklasifikasi *Naive Bayes*, yang terdiri dari dua tahap yang dilakukan, yaitu tahap training dan tahap testing. Untuk menentukan kinerja dari mesin pengklasifikasi *Naive Bayes* dalam proses klasifikasi dilakukan eksperimen menggunakan 10 kali validasi silang (*10-fold cross validation*). Dari hasil eksperimen tersebut didapatkan rata-rata akurasi sebesar 93.33%.

### 4. Desain Penelitian/Metodologi

Jenis penelitian yang dipakai dalam penelitian ini adalah kualitatif, yaitu penelitian yang dilakukan dengan mengumpulkan data dalam bentuk kata-kata atau kalimat dalam sumber informasi yang penulis jadikan objek, dalam hal ini *twitter*. Tahapan yang dilakukan dalam penelitian ini sebagai berikut ; Studi Literatur -> Pengumpulan Data -> Pra-Prosesing Teks->Fitur Seleksi ->Clasification Naive Bayes -> Classified Data. Dalam penelitian ini menggunakan 4 macam data yaitu data tweet, data kata stopword, data kata dasar dan data pengetahuan.

Data tweet yang digunakan pada penelitian ini yaitu tweet pada twitter terhadap calon wakil presiden 2019 dari partai gerindra Sandiaga Salahuddin Uno, pendapat positif, negatif dan netral dari masyarakat di twitter. Pendapat positif yaitu pendapat dimana pengguna twitter memberikan pendapatnya dengan memberikan penilaian secara positif dan mendukung terhadap Sandiaga Salahuddin Uno. Pendapat negatif merupakan pendapat sebaliknya dari pendapat positif yaitu pengguna twitter memberikan pendapatnya dengan memberikan keterangan ketidaksetujuan terhadap calon wakil presiden sandiuno. Sedangkan pendapat netral yaitu pendapat yang tidak menunjukkan mendukung ataupun tidak mendukung terhadap calon wakil presiden Sandiaga Salahuddin Uno. Data tweet didapat dengan memanfaatkan API search twitter dengan mengetikan keyword yang berhubungan dengan tarif dasar listrik. Berikut gambaran tahap – tahap pengumpulan data.



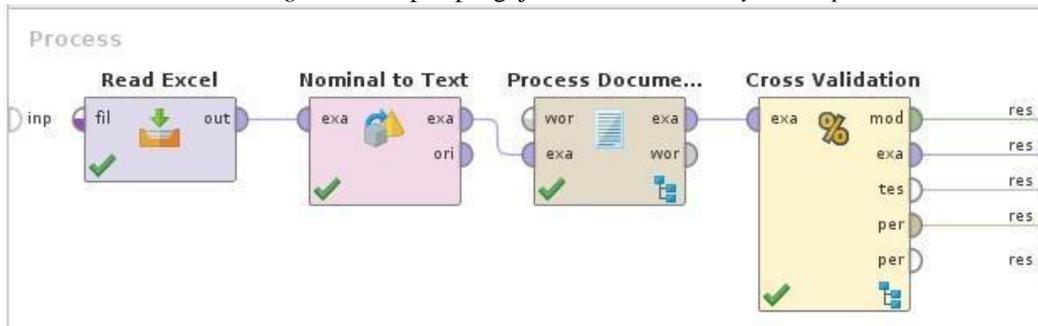
**Gambar 4. Gambar skema pengumpulan data dari *twitter***

Data yang sudah terkumpul tersebut nantinya akan di bagi 2 bagian yaitu untuk data training dan data testing untuk menguji ke akuratan sistem dalam mengklasifikasikan teks.

## 5. Hasil Penelitian

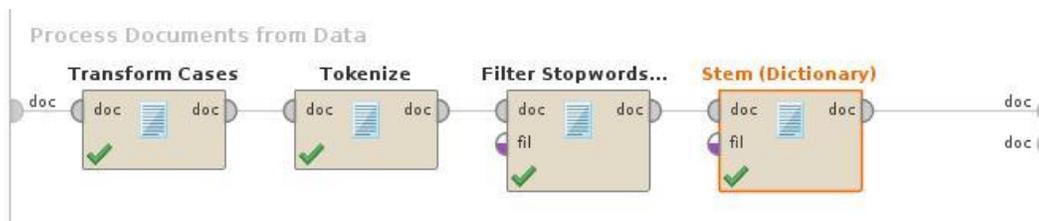
### 5.1. Pengujian Akurasi Metode *Naive Bayes Classifier*

Pengujian akurasi ini dilakukan dengan menggunakan *tool* bantuan yaitu *rapidminer*. Pengujian dilakukan dengan 400 data *tweet* .Berikut *design* dari tahapan pengujian metode *naive bayes* di *rapidminer*.



Gambar 5. Design Pengujian Metode *Naive Bayes*

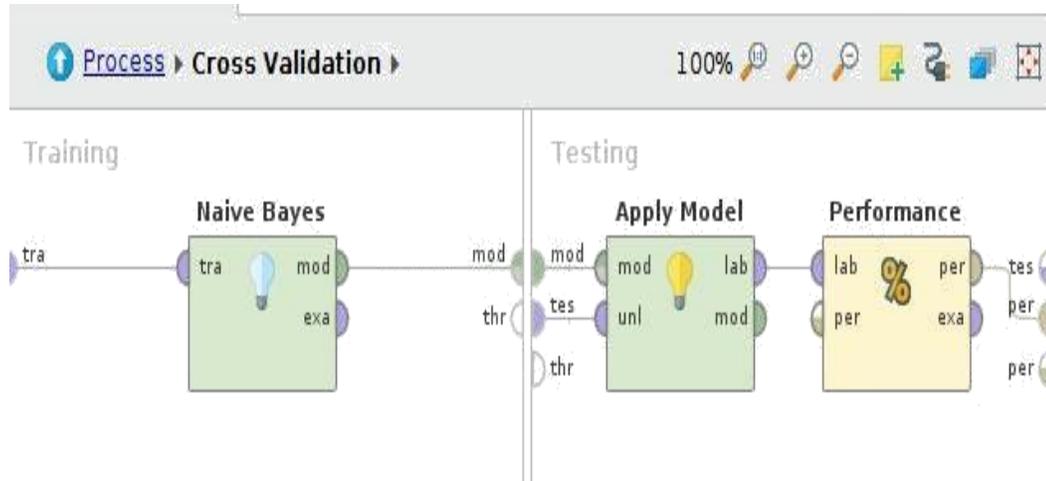
*Read Excel* berfungsi untuk meng-*import dataset* kedalam lingkungan *rapidminer* dalam formal *.xlsx*. Perubahan tipe atribut dari nominal ke teks dengan *operator nominal to text*. *Preprocessing* data dengan *operator process*. Dokumen *from data* yang terdiri dari beberapa proses yang biasa dilakukan pada *preprocessing text* seperti gambar dibawah ini.



Gambar 6. Tahapan *Process Document From Data*

*Transform cases* berfungsi untuk menyeragamkan tulisan ke bentuk *lower case* atau *upper case*. Pembagian kalimat ke bentuk token atau kata perkata tunggal menggunakan *operator Tokenize*. *Filter stopwords* untuk menghapus semua kata *stopword* atau kata yang kurang berpengaruh dalam penentuan sentimen dalam kalimat. Perubahan kata kedalam bentuk *root word* (kata dasar) dengan menggunakan *operator stem*.

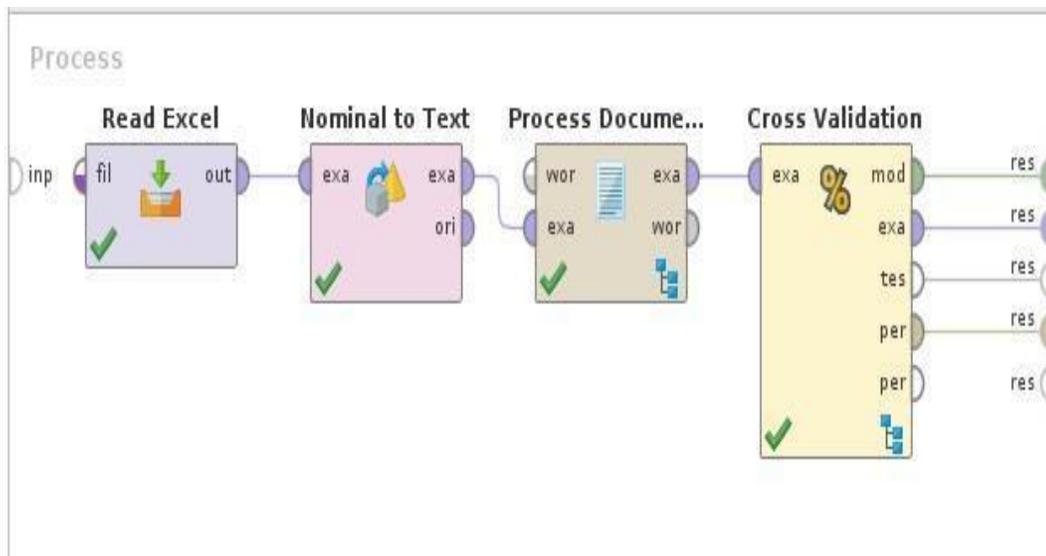
Proses validasi menggunakan *cross validation*. Didalam *cross validation* terdapat dua kolom, *training* dan *testing*. Didalam kolom *training* terdapat algoritma klasifikasi yang diterapkan yaitu *naive bayes*, sedangkan di dalam kolom *testing* terdapat *Apply Model* untuk menjalankan algoritma / *model naive bayes* dan *Performance* untuk mengukur performa dari model *naive bayes* tersebut seperti gambar berikut.



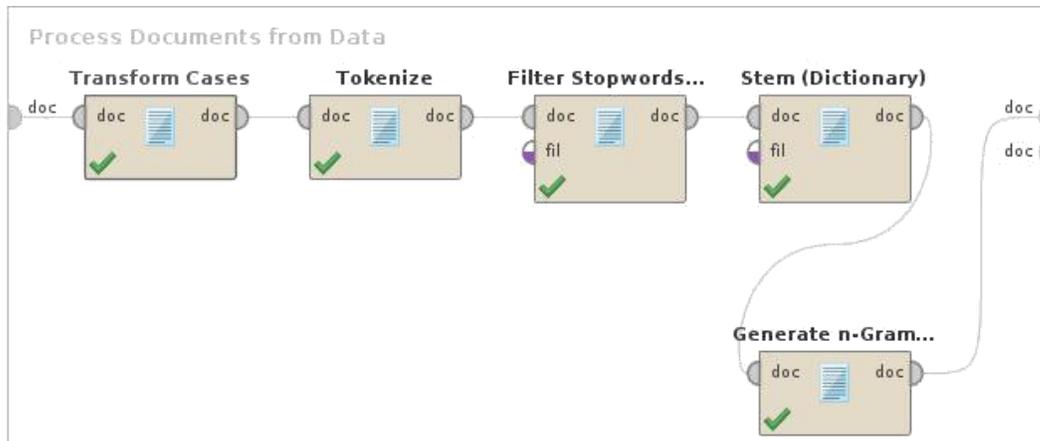
Gambar 7. Validasi Pengujian Naive Bayes

### 5.2. Pengujian Akurasi Metode Naive Bayes Classifier Berbasis N-Gram

Design dari proses pengujian *naive bayes* berbasis *n-gram* pada *rapid miner* hampir sama dengan proses sebelumnya. Berikut *design* proses nya.

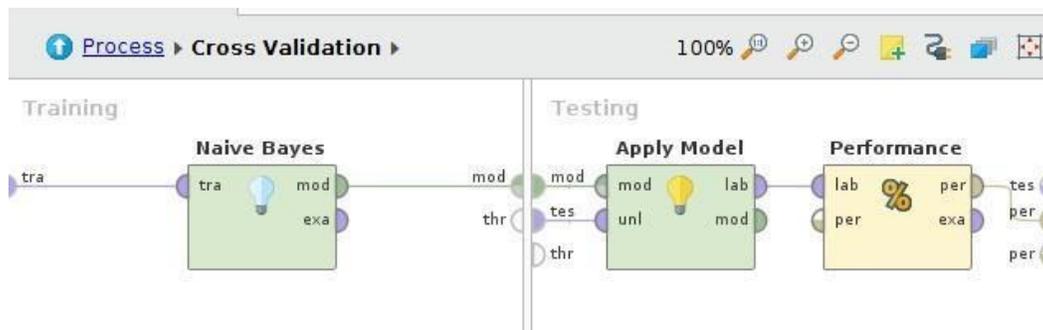


Gambar 8. Design Pengujian Naive Bayes Berbasis N-gram



**Gambar 9.** Tahapan Proses *Text Preprocessing* dengan *N-gram*

Sebelum diterapkan *n-gram*, proses yang sama harus diterapkan terlebih dahulu yaitu *preprocessing text* yang terdiri dari beberapa proses seperti gambar 4.5 diatas yang disimbolkan dengan *operator - operator* yang ada di *rapid miner*. Untuk *n-gram* sendiri menggunakan *operator Generate N-gram (Characters)* berfungsi untuk merubah token dalam kalimat kedalam bentuk *n-gram*. Dalam pengujian ini menggunakan unigram atau membagi kedalam 2 karakter. Kemudian dilanjutkan proses validasi menggunakan *cross validation* seperti gambar berikut

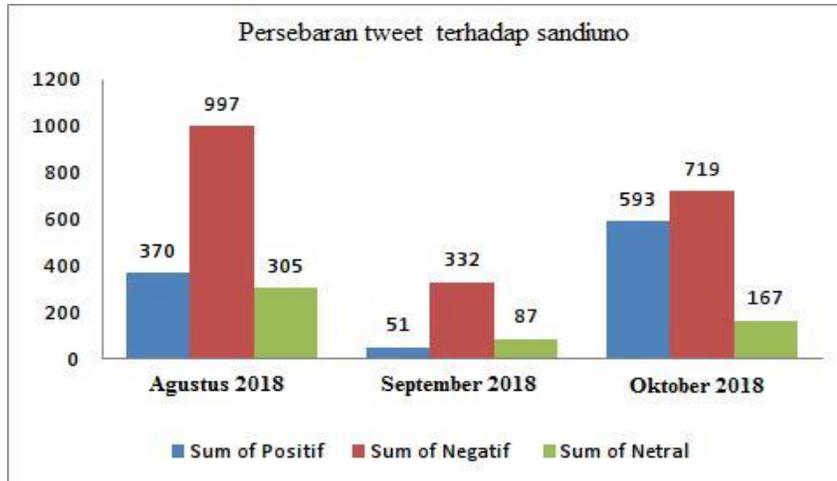


**Gambar 10.** Validasi Penegujian Metode *Naive Bayes* berbasis *N-gram*

Di dalam *cross validation* terdapat dua kolom, training dan testing. Di dalam kolom training terdapat algoritma klasifikasi yang diterapkan yaitu *naive bayes*, sedangkan didalam kolom testing terdapat *Apply Model* untuk menjalankan algoritma / *model naive bayes* dan *performance* untuk mengukur performa dari model *naive bayes* tersebut.

### 5.3. Hasil Analisis Persebaran *Tweet* calon wakil presiden

Dari data yang dikumpulkan dari periode Mei – Juli 2017 sekitar 4000 data *tweet* dari *twitter* yang sudah terklasifikasikan divisualisasikan dalam bentuk grafik seperti berikut:



**Gambar 11.** Grafik Persebaran *Tweet* Periode Agustus-Oktober 2018

Dari grafik diatas bisa kita lihat pada bulan Agustus, *tweet* negatif terhadap calon wakil presiden Sandiaga Salahudin Uno sangat tinggi sedangkan *tweet* negatif paling kecil berada di bulan September.

#### 5.4. Hasil Pengujian Naive Bayes

##### 5.4.1. Metode *Naive Bayes*

Hasil pengolahan 300 data *training* menggunakan algoritma *Naive bayes* pada *rapidminer* menampilkan hasil seperti pada tabel *confusion matrix* berikut.

**Tabel 1.** *Confusion Matrix* Pengujian *Naive Bayes*

Accuracy:89,67% +/-2.37%(mikro:89.67%)				
	<i>True netral</i>	<i>True positif</i>	<i>True negatif</i>	<i>Class precision</i>
<i>Pred.netral</i>	93	0	19	87,25%
<i>Pred.positif</i>	1	100	5	98,04%
<i>Pred negatif</i>	6	0	76	90,62%
<i>Class recal</i>	93,00%	100,00%	76,00%	

Dari gambar diatas diketahui hasil akurasi dari pengujian metode *naive bayes* sebesar 89.67%. Nilai ini termasuk *Good classification*.

##### 5.4.2. Metode *Naive bayes* berbasis *N-gram*

Dari hasil pengujian diketahui bahwa penerapan *n-gram* dapat meningkatkan akurasi dari metode *naive bayes*. Hasil akurasi pengujian menunjukkan angka 92.00% setelah diterapkan *n-gram*. Berikut tabel *confusion matrix* yang ditampilkan *rapidminer*.

**Tabel 2.** Confusion Matrix Pengujian Naïve Bayes Berbasis N-gram

Accuracy:92.00%+/-3.40%(mikro:92.00%)				
	<i>True netral</i>	<i>True positif</i>	<i>True negative</i>	<i>Class precision</i>
<i>Pred.netral</i>	89	0	13	87,25%
<i>Pred.positif</i>	2	100	0	98,04%
<i>Pred negatif</i>	9	0	87	90,62%
<i>Class recal</i>	89,00%	100,00%	87,00%	

Setelah melakukan pemodelan dan perhitungan berdasar kedua algoritma diatas, kemudian dilakukan perbandingan hasil yang berupa nilai akurasi. Maka diperoleh data perbandingan sebagai berikut :

**Tabel 3.** Perbandingan Hasil Pengujian

Perbandingan	Naive Bayes	Naive Bayes Berbasis N-Gram
Accuracy	89.67%	92.00%

Tabel diatas merupakan hasil akhir dari pengujian yang menunjukan perbandingan tingkat akurasi dari metode *naïve bayes* tanpa ekstraksi fitur *n-gram* dan metode *naïve bayes* dengan ditambah ekstraksi fitur *n-gram*. Berdasarkan table diatas diketahui bahwa penambahan ekstraksi fitur *n-gram* memberikan peningkatan nilai akurasi terhadap metode *naïve bayes* dalam mengklasifikasikan teks sekitar 2.33%.

## 6. Kesimpulan

Berdasarkan dari hasil penelitian terhadap model algoritma *naïve bayes* tingkat akurasi dari metode *naïve bayes* pengaruh ekstraksi fitur *n-gram* terhadap nilai akurasi dari metode *naïve bayes* dalam mengklasifikasikan data *tweet* adalah:

Pada penelitian ini algoritma *naïve bayes* mampu mencapai tingkat akurasi 89,67%, dan pengaruh ekstraksi fitur *n-gram* yang diterapkan dapat meningkatkan nilai akurasi pada algoritma *naïve bayes* sekitar 2,33%, yaitu menjadi 92,00% dalam pengklasifikasian data *tweet* yang diukur dengan *rapidminer*.

Diharapkan penelitian ini berguna untuk mengetahui tingkat sentiment masyarakat pada media social lainnya terhadap suatu tokoh atau tren terbaru kedepannya akan ada pengembangan sentiment analisis dengan metode yang berbeda sehingga dapat dibandingkan dengan metode *naïve bayes* dengan *rapidminer*

## Daftar Pustaka

- Nugroho, E., 2011, Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karpi, *Skripsi*, Program Studi Ilmu Komputer, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya Malang.
- Romelta, E., 2012, Opinion Mining di Twitter untuk Customer Feedback Smartphone dengan Pembelajaran Mesin, *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, Vol. 1, No. 2, Juli 2012.
- Sang, E. T. K., dan Bos, J., 2012, Predicting The 2011 Dutch Senate Election Results with Twitter, *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 53-60.
- Sunni, I., dan Widyantoro, D.H., 2012, Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini *SIGMA – Jurnal Teknologi Pelita Bangsa*

terhadap Tokoh Publik, *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, Vol. 1, No. 2, Juli 2012.

Kotwal, Aishwarya et al. 2016. *Improvement in Sentiment Analysis of Twitter Data using Hadoop*. International Conference on “Computing for Sustainable Global Development”, 16th – 18th March, 2016. BVICAM, New Delhi (INDIA).

Kumar, Lokesh and Bhatia, Parul Kalra. 2015. *Text Minig: Concepts, Process And Application*. Journal of Global Research in Computer Science. Volume 4, No. 3, March 2013