

## Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes

Aprilia Wandani<sup>1</sup>, Fauziah<sup>2</sup>, Andrianingsih<sup>3</sup>

Sistem Informasi, Fakultas Teknologi Komunikasi dan Informasi  
Universitas Nasional Jl. Sawo Manila, Pejaten Ps. Minggu Jakarta 12520, 021-  
7806700

aprilwan411@gmail.com, fauziah@civitas.unas.ac.id, andrianingsih@civitas.unas.ac.id

### Abstract

*There is a sales system called Flash Sale in e-commerce. Basically the concept of a Flash Sale is to offer a lower price and a predetermined time and number of products. The sales system is only held at certain moments, by making cheaper product sales but with a limited time and number of products it will make sales increase because buyer interest will be higher. But apart from all the advantages of course there will be pros and cons. Sentiment analysis on Twitter was chosen because Twitter itself is a social media that allows users to be free to comment or write opinions about anything, including opinions about flash sale events that exist in e-commerce today. Thus, this research exists to find out the opinions of existing Twitter users regarding the Flash Sale event held by e-commerce. By using the methodology of three classification algorithms, Naive Bayes, K-Nearest Neighbor and Random Forest in classifying the data to determine the accuracy of the sentiment value of Twitter users in the Flash Sale event. This research takes two data samples from the keywords "flash sale" and "flash sale shopee", the results accuracy of the implementation of the three classification algorithms are 83.53% Naive Bayes, 82.94% K-NN, 80.59% Random Forest for the keyword "flash sale" and 81.48% Naive Bayes, 77.78% K-NN, 74.07% Random Forest for the keyword "flash sale shopee". With this, the Naive Bayes Algorithm becomes a recommendation for classifying Sentiment Analysis data with greater accuracy and more stability to be used for large and small data.*

**Keywords:** Sentiment Analysis, Flashsale, Naive Bayes, k-NN, Random Forest

### Abstrak

*Dalam dunia e-commerce terdapat sistem penjualan yang disebut Flash Sale. Pada dasarnya konsep dari Flash Sale yaitu memberikan tawaran harga yang lebih rendah dan telah di tentukan waktu dan jumlah produk. Sistem penjualan hanya diadakan dalam momen tertentu saja, dengan membuat penjualan produk yang lebih murah tetapi dengan waktu dan jumlah produk yang terbatas maka akan membuat penjualan lebih meningkat di karenakan minat pembeli yang akan lebih tinggi. Namun lepas dari segala kelebihan tentu saja akan ada pro dan kontra. Analisis sentimen pada twitter ini dipilih karena twitter sendiri merupakan salah satu media sosial yang menyediakan kebebasan untuk berkomentar atau menulis opini tentang apapun termasuk opini tentang event flash sale yang ada pada e-commerce saat ini. Dengan demikian, penelitian ini ada untuk mengetahui opini pada pengguna Twitter yang ada berkaitan dengan event Flash Sale yang diadakan e-commerce. Dengan menggunakan metodologi tiga algoritma klasifikasi yaitu Naive Bayes, K-Nearest Neighbour dan Random Forest dalam pengklasifikasian data untuk mengetahui akurasi tingkat nilai sentimen pengguna Twitter pada event Flash Sale. Penelitian ini mengambil dua sampel data dari kata kunci "flash sale" dan "flash sale shopee", hasil dari implementasi ketiga algoritma klasifikasi tersebut yaitu akurasi sebesar 83.53% Naive Bayes, 82.94% K-NN, 80.59% Random Forest untuk kata kunci "flash sale" dan 81.48% Naive Bayes, 77,78% K-NN, 74.07% Random Forest untuk kata kunci "flash sale shopee". Dengan ini Algoritma Naive Bayes menjadi rekomendasi*

untuk pengklasifikasian data Analisis Sentimen dengan akurasi lebih besar dan lebih stabil digunakan untuk data yang besar maupun kecil.

**Kata kunci:** Analisis Sentimen, Flash Sale, Naive Bayes, k-NN, Random Forest

## 1. PENDAHULUAN

Saat ini masyarakat modern sangat bergantung pada penggunaan teknologi, termasuk untuk berbelanja. Masyarakat tidak perlu untuk pergi keluar rumah untuk membeli barang-barang yang dibutuhkan, cukup dengan mencari di *e-commerce* yang telah banyak tersedia saat ini. Pada *e-commerce* terdapat sistem penjualan yang disebut Flash Sale. Sistem ini mempunyai konsep menjual dengan harga yang murah dalam waktu dan jumlah produk yang ditentukan. Dengan terbatasnya waktu dan jumlah produk ini membuat produk terjual dengan cepat dan dikarenakan harga yang lebih rendah peminatnya pun semakin meningkat. Selain dari segala keuntungannya, momen Flash Sale ini pasti tidak lepas dengan adanya pro dan kontra. Melalui media sosial twitter, kita dapat mencari informasi tentang tingkat pro dan kontra yang berkaitan dengan Flash Sale pada *e-commerce*. Twitter merupakan salah satu media sosial yang populer dikarenakan pada media sosial ini kita dapat dengan bebas berkomentar atau beropini mengenai apapun[1]. Dari sekian banyak opini yang ada pada twitter ini pasti ada yang bersifat positif maupun negatif. Dalam penelitian ini, tingkat akurasi pada perbandingan antara penggunaan algoritma klasifikasi akan menjadi proses dalam memprediksi bagaimana tingkat penilaian pengguna twitter pada event flash sale.

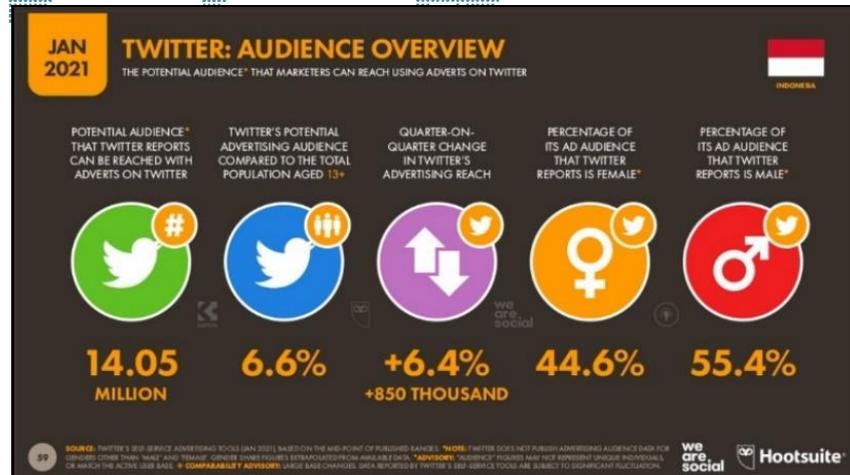
Pendapat atau komentar merupakan salah satu acuan penilaian terhadap suatu objek atau kejadian. Hasilnya kita dapat mengetahui langkah-langkah selanjutnya untuk memperbaiki apa saja kekurangan atau kelebihan yang telah kita ketahui dari berbagai opini yang ada[2]. Pada penelitian tugas akhir ini adapun masalah yang diangkat adalah analisis sentimen masyarakat terhadap event Flash Sale yang ada pada *e-commerce*. Data analisis yang digunakan berdasarkan komentar yang ada pada twitter dengan kata kunci "flash sale" dan "flash sale shopee" bertujuan untuk mengetahui penilaian pengguna Twitter antara event Flash Sale secara umum dan secara spesifik dengan merujuk *e-commerce* Shopee. Tujuan ini bisa digunakan untuk penilaian terhadap Flash Sale yang merupakan salah satu sistem penjualan yang telah diadakan oleh sebagian besar *e-commerce*. Serta untuk melihat tingkat akurasi algoritma klasifikasi dalam memprediksi Sentimen Analisis.

## 2. METODOLOGI PENELITIAN

### 2.1. Data Twitter

Jumlah pengguna media sosial saat ini terus meningkat, menurut Hootsuite dan We Are Social ampai dengan Januari 2021 ada 170 juta pengguna media sosial di Indonesia. Jumlah pengguna media sosial di

Indonesia bertambah 10 juta dari tahun 2020, dan Twitter menduduki peringkat 5 teratas media sosial yang paling sering digunakan di Indonesia.



**Gambar 1.** Data Pengguna Twitter Di Indonesia Januari 2021  
(Di ambil dari [www.wearesocial.com](http://www.wearesocial.com))

Pada gambar 1 dapat dilihat terdapat data 850ribu pengguna Twitter di Indonesia dari Januari 2021 dengan persentase 44.6% merupakan pengguna perempuan dan 55.4% pengguna laki-laki dengan hastag yang mencapai 14.5 juta. Twitter menyimpan beberapa ciri yang membuatnya menjadi data mining yang sangat menarik. Walaupun ada kendala berbagai karakter didalamnya, data mining menghadirkan kinerja yang lebih rendah daripada teks mining yang lebih panjang. Selain itu, partisi menjadi beberapa kelas tetap menjadi tantangan berat: kelas biner konten tekstual biasanya bergantung pada polaritas sentimen dari aditifnya sementara ketika pelatihan positif dan negatif dibagi menjadi subkelas, akurasinya cenderung sangat rendah[3].

Pada media sosial Twitter kita dapat berkomentar atau pesan yang dapat dibaca oleh Twitter *User* lainnya, fitur ini biasa kita sebut tweet atau kicauan. Pada Twitter kita bisa membuat tweet dengan maksimum 140 karakter yang membuat media sosial tersebut unik. Serta API (*Application Programming Interface*) yang disediakan Twitter sangat baik sehingga kelebihan dari twitter ini memudahkan kita mengambil data dari Twitter[4]. Sebagai sumber informasi, Komponen twitter yang dapat digunakan untuk pengambilan informasi :

- User name : identifikasi pengguna
- Hashtags : tagar untuk mengaitkan topik tertentu
- Time Stamp : waktu saat tweet itu dikirim
- Replies : balasan pesan
- Tweet Text : isi dari tweet
- Retweet : membagikan tweet

Data Twitter dapat diakses dengan API REST Twitter yang telah disediakan oleh pihak Twitter dengan mengajukan permintaan pada pihak Twitter untuk mendapatkan API Key. Twitter dipilih karena Twitter sebagian besar cocok untuk data mining karena tiga point utama[5] :

- a) Twitter API memiliki desain yang bagus dan mudah di akses.
- b) Data Twitter dalam format yang nyaman untuk dianalisis.
- c) Kebijakan Twitter untuk data relatif liberal dibandingkan dengan API yang lain.

Keterbatasan platform media sosial lain seperti Facebook adalah mereka tidak mengizinkan akses mudah ke data mereka karena kebijakan privasi yang berbeda-beda. Tidak mudah mengumpulkan data secara terbuka dan otomatis dengan media lain semacam itu. Opini dari sebuah tweet dapat ditemukan di dalam bagian teks dari tweet tersebut. Penarikan data dalam bentuk teks bebas yang tidak terstruktur, tidak berstandar[6].

## 2.2. *Sentiment Analysis*

Analisis Sentimen adalah sebuah proses untuk menentukan atau mengukur nilai sentimen atau opini yang ada terhadap suatu objek atau kejadian yang berupa teks dan dapat dikategorikan sebagai sentimen positif, negatif, serta netral[7]. Pengguna internet saat ini banyak yang menuliskan pendapat atau opini, pengalaman dan berbagai hal yang terjadi atau sekiranya menarik untuk mereka[8].

Analisis sentimen dianggap sebagai masalah pengelompokan. Sama seperti dalam laporan besar, nilai sentimen pada tweet dapat dikomunikasikan dalam berbagai cara dan ditandai dengan adanya sentimen didalamnya. jika ada sentimen dalam tweet, mengandung *polar word* atau kata-kata berlawanan maka itu ditetapkan positif atau negatif, jika tidak dianggap Netral[4]. Langkah analisis sentimen sebagai berikut[9]:

- a) Level 1: Mencari sentimen negatif, positif, serta netral pada setiap baris.
- b) Level 2: Analisa sentimen seluruh dokumen sebagai negatif, positif, serta netral.
- c) Level 3: Menerapkan pengelompokan dimana mengumpulkan semua atribut yang ada dengan hasil sentimen yang sama.
- d) Level 4: Memanfaatkan visualisasi data dari analisis sentiment untuk interaksi antar *user*.

## 2.3. *Text Mining*

*Text mining* bertujuan untuk menganalisis opini, sentiment, evaluasi, penilaian, sikap serta emosi seseorang sehingga dapat diketahui kaitannya dengan suatu topik, layanan, organisasi, individu, atau kegiatan tertentu. *Text mining* dan *data mining* berbeda dalam artian melibatkan data yang terstruktur untuk *data mining* sedangkan untuk *text mining* erat kaitannya dengan fitur tertentu dan data tidak terstruktur sehingga memerlukan

preprocessing untuk pengolahan data. Meskipun demikian keduanya memiliki konsep yang dalam persepsi ilmu algoritma yang sama. Berikut proses *text mining*[10]:

- a) Ekstraksi Informasi: dimana langkah awal analisis teks yang tidak terstruktur dengan proses penguraian data , informasi data dan hubungan frasa untuk data yang besar.
- b) Clustering: Teks yang ada dipisahkan menjadi satu kelompok dengan cara memberi nilai, daftar kelas yang akan diberi nilai berdasarkan algoritma clustering untuk menghitung persamaan atau *similarity*.
- c) Klasifikasi: Menemukan informasi utama dengan menentukan data utama berupa pelabelan kelas untuk analisis data yang ada.
- d) Visualisasi data: Menampilkan hasil ekstraksi data yang akan digunakan pengguna untuk interaksi dengan dokumen yang ada.

#### 2.4. Klasifikasi

Dalam metode yang paling umum digunakan pada data mining ini dilakukan tindakan pengelompokan pada setiap keadaan. Setiap keadaan yang akan berisikan sekelompok atribut yang disebut class attribute. Klasifikasi membantu menemukan label kelas yang sesuai untuk digunakan pada *supervised learning* yang membutuhkan kumpulan data yang diberi label dengan baik untuk *training data*[11]. Ada 4 komponen dasar yang dimiliki Klasifikasi, yaitu [12]:

- a) *Class* : Variabel yang akan menjadi label.
- b) *Predictor* : Variabel yang akan menjadi atribut dari data yang ada untuk digunakan dalam pengklasifian.
- c) *Training dataset* : Data yang telah memiliki label sebelumnya.
- d) *Testing dataset* : Data yang akan dilakukan proses klasifikasi.

#### 2.5. Crawling

*Crawling* merupakan tahapan yang penting untuk dilakukan saat akan melakukan analisis data, pada teknik pengumpulan data ini melakukan proses pengumpulan data Twitter dengan memanfaatkan *Application Programming Interface* (API) yang disediakan oleh Twitter yang menghasilkan kumpulan data sampel berupa teks yang telah diunggah oleh pengguna Twitter[13].

Mengambil informasi dari Twitter menggunakan operator "*Search Twitter*" dan menyimpan informasi dalam catatan dominan menggunakan operator "*Write Excel*" dengan hanya memisahkan data dari tweet dan menghapus semua salinan tweet dengan konten "*Remove Duplicate*". Dan selanjutnya melakukan *labeling*. membuat partisi data informasi menjadi beberapa kelas penilaian yang akan dimanfaatkan pada tahap selanjutnya. Kuantitas kelas opini yang digunakan ada tiga kelas, yaitu spesifik negatif, positif, serta netral. Alasan siklus penandaan ini adalah untuk memisahkan

dataset menjadi 2 bagian, menjadi informasi penyiapan dan informasi pengujian testing[14].

## 2.6. Naive Bayes Classifier

Naive Bayes Classifier merupakan salah satu metode pengklasifikasian data yang bersumber pada teorema Bayes. Metode yang dicetuskan oleh ilmuwan inggris Thomas Bayes ini menggunakan metode probabilitas dan statistik. Pada metode ini dilakukan 2 tahapan proses klarifikasi teks, tahapan klasifikasi dan tahapan *training*. Proses analisis data yang ada dengan pemilihan kata yang mungkin muncul menjadi representasi dokumen merupakan tahap *training*, sedangkan pada tahap klasifikasi pada suatu dokumen ditentukan nilai kategori berdasarkan term yang akan muncul dalam pengklasifikasian dokumen[15]. Berikut bentuk umum dari Naive Bayes Classifier [16]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Dimana X= data kelas asing; H = kelas khusus (Hipotesa data X);  $P(H|X)$  = prediksi hipotesis H pada X;  $P(H)$  = prediksi hipotesis H;  $P(X|H)$  = prediksi hipotesis X pada H; dan  $P(X)$  = prediksi X.

## 2.7. Random Forest

Random Forest dilakukan untuk mendapatkan hasil akhir dari sistem identifikasi, sebuah prediksi akan diperoleh dengan mengamati *voting* mayoritas dari setiap kelas yang telah dilakukan pelabelan data sebelumnya[17]. Random Forest merupakan salah satu algoritma klasifikasi yang dilakukan dengan penggabungan tree dengan training pada data yang dimiliki. Pada Random Forest di lakukan proses pengklasifikasian dengan memecah data secara acak ke dalam decision tree.

## 2.8. K-NN (K-Nearest Neighbor)

KNN adalah jenis berbasis *instance learning* atau *lazy learning*. Algoritma ini yang paling sederhana dari semua algoritma *machine learning*[18]. Merupakan salah satu dari algoritma klasifikasi yang memiliki konsep melihat jarak tetangga terdekat dengan data baru. Algoritma ini akan mencari jarak yang paling dekat antara *training set* dengan k tetangga terdekat dari *testing set*. Algoritma K-Nearest Neighbor memiliki konsep menggolongkan dokumen uji X sebagai berikut[19]:

- a) Menetapkan Parameter K yang merupakan jumlah dari tetangga terdekat
- b) Pada data training dilakukan pengkalkulasian kuadrat jarak euclidean
- c) Mengurutkannya secara ascending
- d) Menggabungkan kategori Y yang berlandaskan tetapan parameter K

## 2.9. Rapid Miner

Perangkat lunak ini bersifat open source, dengan Rapid Miner dapat menganalisis *data mining*, *text mining*, serta analisis prediksi. Aplikasi ini juga mendukung *machine learning* seperti *preparing data*, *processing*, *modeling*, hingga visualisasi. Alat ini menawarkan lebih dari 1000 operator drag-and-drop yang dapat digunakan untuk melakukan operasi data mining, dengan mudah dan cepat. Sering digunakan untuk tujuan bisnis dan komersial dan juga untuk pendidikan, penelitian, pelatihan, pembuatan prototipe cepat, dan pengembangan aplikasi dan *model deployment*[20]. Pada penelitian ini semua proses dari proses awal *crawling* data hingga visualisasi dilakukan dengan Rapid Miner.

## 3. HASIL DAN PEMBAHASAN

Pada gambar 2 merupakan kerangka penelitian berupa langkah-langkah dalam proses penelitian ini dari awal pengumpulan atau penarikan data (*crawling*) dari Twitter sampai visualisasi.

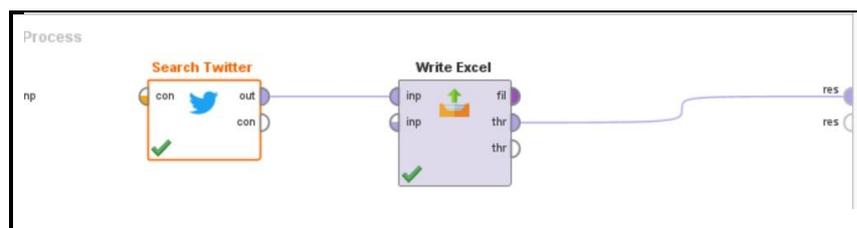


Gambar 2. Kerangka Penelitian

### 3.1. Pengumpulan Data

Pada tahapan awal yaitu proses pengumpulan data yang ada pada media sosial twitter. Data yang digunakan adalah komentar pengguna media sosial Twitter yang ada pada kata kunci "flash sale" dan "flash sale shopee" dengan filter bahasa indonesia. Data yang diperoleh yaitu *twit* dari pengguna Twitter dalam 7 hari kebelakang, hal ini adalah kebijakan dari API Twitter.

- Crawling* : Data diperoleh secara langsung terhubung dari media sosial Twitter dengan menggunakan Rapid Miner
- Menyimpan Data: Data yang telah didapatkan disimpan dalam bentuk excel.



Gambar 3. Proses Crawling Data Twitter

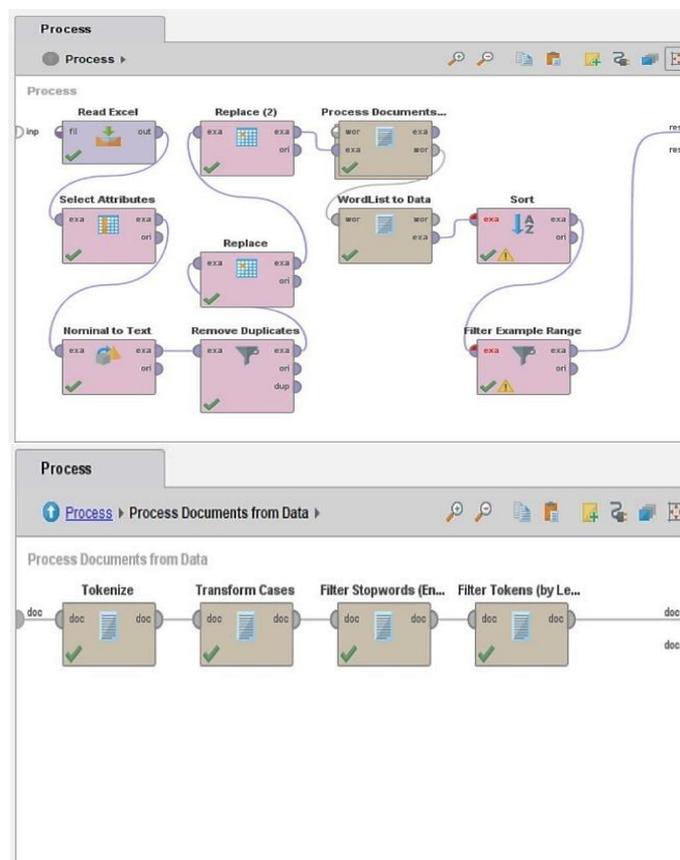
Pada gambar 3 digambarkan proses *crawling* dengan properties "Search Twitter" dengan *filtering* menggunakan bahasa indonesia yang sebelumnya telah melakukan koneksi pada akun Twitter lalu data yang

diperoleh disimpan dalam bentuk excel. Pada tahap ini diperoleh 868 sampel data awal untuk kata kunci “flash sale” dan 149 sampel data awal untuk kata kunci “flash sale shopee”.

### 3.2. Pengolahan Data(Preprocessing)

Pada tahapan ini bertujuan untuk mengekstrak dan membersihkan data yang telah di peroleh dari crawling sebelumnya. Adapun tahapan dari *Preprocessing* ini adalah:

- Penghapusan Data Berulang (Remove Duplicate)
- Pemilihan Atribut
- Penghapusan tanda baca (Replace)
- Filterisasi stopwords
- Tokenisasi
- Mencari frekuensi kata-kata yang sering muncul

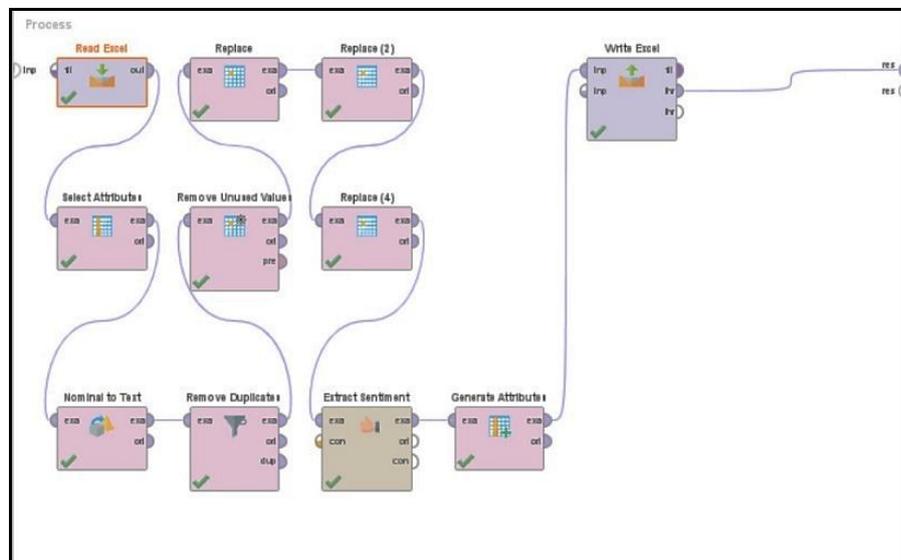


**Gambar 4.** Proses Pengolahan Data

Pada gambar 4 data yang telah diperoleh sebelumnya dioalah untuk membersihkan data yang ada agar dapat diperoleh data *Term frequency* atau kata apa saja yang sering muncul.

### 3.3. Analisis Sentimen

Pada tahap ini data diolah agar dapat dilakukan labeling dan penambahan atribute untuk mengetahui nilai sentimen dari data yang berupa negatif , positif atau netral.



**Gambar 5.** Proses Analisis Sentimen

Pada gambar 5 merupakan proses Analisis Sentimen, pada tahap ini dilakukan pembersihan data, namun hanya dengan menghapus data berulang (*Removing Duplicate*), dan menghapus tanda baca serta values yang tidak digunakan (*Replace*) pada attribute *text* yang sudah dipilih (*Select Attribute*). Karena data yang ingin dinilai adalah data dari setiap komentar pada pengguna Twitter dalam bentuk kalimat. Setelah data diolah, selanjutnya proses Analisis Sentimen dengan *Extract Sentiment* menggunakan model Vedar dan akan di peroleh nilai-nilai yang selanjutnya data akan dilakukan *labeling* dengan menambahkan attribute Sentimen dengan *Generate Attribute* sehingga menghasilkan nilai sentimen positif, negatif, serta netral berdasarkan data nilai yang diperoleh dari *Extract Sentiment* yang selanjutnya disimpan dalam bentuk excel.

**Tabel 1.** Data yang diperoleh setelah proses Analisis Sentimen

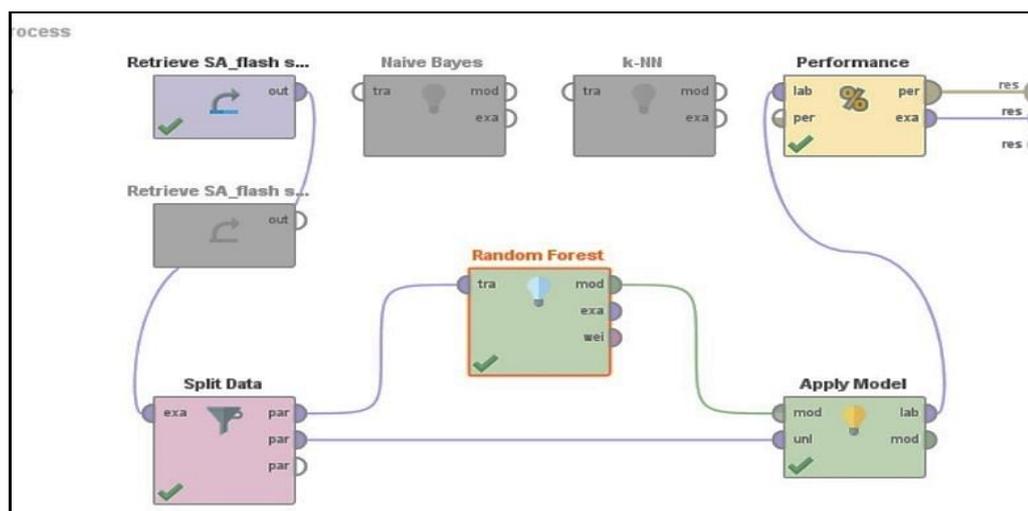
Kata Kunci	Data	Negatif	Positif	Netral
flash sale	567	32 (5.6%)	85 (14.9%)	450 (79.3%)
flash sale shopee	91	7 (7.7%)	15 (16%)	69 (75.8%)

Pada tabel 1 diperoleh data hasil proses Analisis Sentimen, data yang ada berbeda dari data awal karena proses preprocessing sebelumnya lalu data telah dilakukan pelabelan sehingga diperoleh nilai sentimen negatif sebesar 32(5.6%), positif 85(14.9%), netral 450(79.3%) dari keseluruhan data sebesar 567 data sampel dengan kata kunci "flash sale", sedangkan

untuk data dengan kata kunci “flash sale shopee” diperoleh negatif 7 (7.7%), Positif 15(16%), 69(75.8%) dari keseluruhan data sampel sebesar 91.

### 3.4. Implementasi Metode Klasifikasi

Setelah melakukan Analisis Sentimen tahap selanjutnya adalah pengolahan data yang telah *labeling* untuk selanjutnya dilakukan pengklasifikasian. Metode yang digunakan untuk pengklasifikasian dalam penelitian ini yaitu menggunakan Metode Naive Bayes Classifier, K-Nearest Neighbour, serta Random Forest menggunakan Rapid Miner.



**Gambar 6.** Proses Implementasi Algoritma Klasifikasi

Pada gambar 7 memperlihatkan proses implementasi data ke dalam Algoritma K-NN, Random Forest dan Naive Bayes. Data yang sebelumnya telah dioalah disimpan dalam bentuk excel lalu dilakukan *split data* untuk menentukan data *training* dan data *testing* dengan ratio 7:3 serta menggunakan tipe *Shuffled Sampling*, lalu diproses dengan properties *Performance* untuk mengetahui besar akurasi dari setiap algoritma.

	true neutral	true positive	true negative	class precision
pred. neutral	137	25	3	83.03%
pred. positive	0	0	0	0.00%
pred. negative	0	0	5	100.00%
class recall	100.00%	0.00%	62.50%	

**Gambar 7.** Akurasi Naive Bayes pada kata kunci "flash sale"

The screenshot shows a performance vector for a Naive Bayes classifier. The accuracy is 81.48%. The confusion matrix is as follows:

	true neutral	true positive	true negative	class precision
pred. neutral	21	4	1	80.77%
pred. positive	0	0	0	0.00%
pred. negative	0	0	1	100.00%
class recall	100.00%	0.00%	50.00%	

**Gambar 8.** Akurasi Naive Bayes pada kata kunci "flash sale shopee"

Pada gambar 8 dan 9 dapat dilihat besar akurasi Naive Bayes 83.53% untuk data kata kunci "flash sale" dan 81.48% untuk data kata kunci "flash sale shopee".

The screenshot shows a performance vector for a K-NN classifier. The accuracy is 82.94%. The confusion matrix is as follows:

	true neutral	true positive	true negative	class precision
pred. neutral	137	25	4	82.53%
pred. positive	0	0	0	0.00%
pred. negative	0	0	4	100.00%
class recall	100.00%	0.00%	50.00%	

**Gambar 9.** Akurasi K-NN pada kata kunci "flash sale"

The screenshot shows a performance vector for a K-NN classifier. The accuracy is 77.78%. The confusion matrix is as follows:

	true neutral	true positive	true negative	class precision
pred. neutral	21	4	2	77.78%
pred. positive	0	0	0	0.00%
pred. negative	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

**Gambar 10.** Akurasi K-NN pada kata kunci "flash sale shopee"

Pada gambar 10 dan 11 dapat dilihat besar akurasi K-Nearest Neighbour sebesar 82.94% untuk data kata kunci "flash sale" dan 77.78% untuk data kata kunci "flash sale shopee".

Random Forest Model (Random Forest) PerformanceVector (Performance)

Table View Plot View

accuracy: 80.59%

	true neutral	true positive	true negative	class precision
pred. neutral	137	25	8	80.59%
pred. positive	0	0	0	0.00%
pred. negative	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

**Gambar 11.** Akurasi Random Forest pada kata kunci "flash sale"

Random Forest Model (Random Forest) PerformanceVector (Performance)

Table View Plot View

accuracy: 74.07%

	true neutral	true positive	true negative	class precision
pred. neutral	20	6	1	74.07%
pred. positive	0	0	0	0.00%
pred. negative	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

**Gambar 12.** Akurasi Random Forest pada kata kunci "flash sale shopee"

Pada gambar 12 dan 13 dapat dilihat besar akurasi Random Forest 80.59% untuk data kata kunci "flash sale" dan 74.07% untuk data kata kunci "flash sale shopee". Dari proses implementasi Algoritma sebelumnya diperoleh data sebagai berikut:

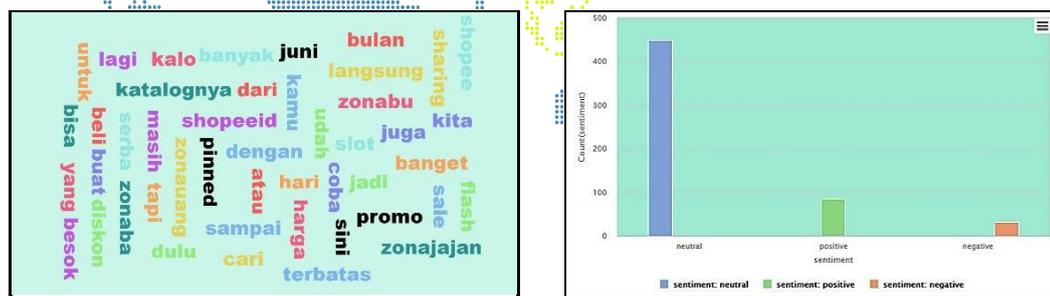
**Tabel 2.** Tabel akurasi dari implementasi Algoritma Klasifikasi

Kata Kunci	Naive Bayes	K-NN	Random Forest
flash sale	83.53%	82.94%	80.59%
flash sale shopee	81.48%	77.78%	74.07%

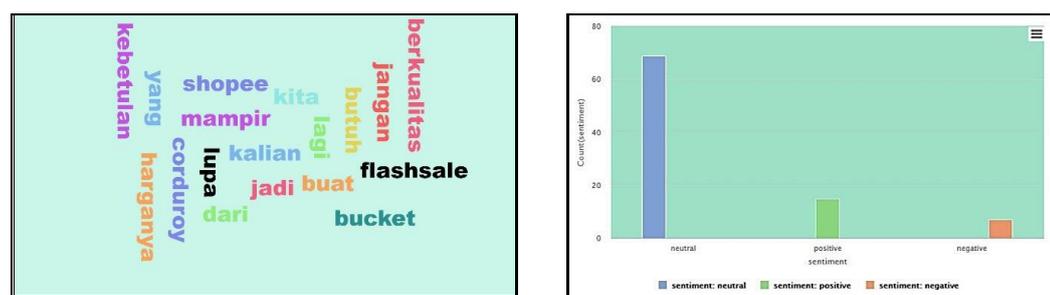
Pada tabel 2 dapat dilihat komparasi besaran akurasi antara algoritma Naive Bayes, K-NN serta Random Forest dari dua data sampel dengan kata kunci "flash sale" dan "flash sale shopee". Algoritma Naive Bayes terlihat lebih baik dalam akurasi dan lebih stabil dalam pengklasifikasikan data besar maupun data kecil.

### 3.5. Visualisasi

Tahap akhir yaitu memvisualisasikan hasil dari analisis sentimen menggunakan *wordcloud* pada Rapid Miner. Dengan menggunakan *wordcloud* dapat terlihat gambaran frekuensi kata-kata yang sering muncul.



Gambar 13. Wordcloud & (Bar Column) "flash sale"



Gambar 14. Wordcloud & Bar (Column) "flash sale shopee"

Pada gambar 15 dan 16 merupakan tampilan visualisasi dari data yang telah diolah, dengan *wordcloud* kita dapat melihat frekuensi kata apa saja yang sering muncul sedangkan Bar(Column) terdapat visualisasi perbandingan antara nilai sentimen netral, positif maupun negatif.

#### 4. SIMPULAN

Berdasarkan dari seluruh proses Analisis Sentimental dengan menggunakan aplikasi Rapid Miner dalam menentukan tingkat penilaian pengguna Twitter terhadap event Flash Sale, dapat disimpulkan bahwa :

- 1) Tanggapan terbanyak adalah Netral, yang berarti mayoritas pengguna Twitter tidak terlalu antusias pada event Flash Sale ini. Akan tetapi tingkat Positif masih lebih banyak dari pada Negatif, yang berarti event Flash Sale masih dinilai baik bagi sebagian orang.
- 2) Berdasarkan frekuensi kata yang muncul pada kata kunci "flash sale" terdapat kata "shopee" walaupun tidak mencari kata tersebut yang menunjukkan bahwa event Flash Sale ini lebih banyak penggemarnya pada *e-commerce* Shopee.
- 3) Penggunaan 2 penarikan data dari kata kunci "flash sale" dan "flash sale shopee" bertujuan untuk mengetahui perbedaan nilai event Flash Sale secara umum dan secara khusus memiliki persentase yang sedikit berbeda. Untuk event Flash Sale yang diadakan oleh Shopee lebih besar tingkat persentasenya dibandingkan event Flash Sale secara umum.

- 4) Untuk implementasi Algoritma klasifikasi menggunakan 3 metode Naive Bayes, K-NN, serta Random Forest diatas dapat dilihat bahwa metode Naive Bayes lebih besar akurasi pada kedua data sampel yang ada yaitu 83.53% untuk kata kunci “flash sale” dan 81.48% pada kata kunci “flash sale shopee”. Maka dapat disimpulkan metode Naive Bayes dapat menghasilkan akurasi lebih baik dalam klasifikasi Analisis Sentimen ini.
- 5) Untuk metode K-NN dan Random Forest memiliki persentase yang cukup baik untuk kata kunci “flash sale” keduanya menghasilkan akurasi lebih dari 80% (82.94% untuk K-NN, 80.59% untuk Random Forest), akan tetapi pada kata kunci “flash sale shopee” akurasi nya menurun menjadi dibawah 80% (77.78% untuk K-NN, 74.07% untuk Random Forest). Hal ini menunjukkan kinerja algoritma K-NN dan Random Forest kurang baik dalam data sampel kecil, berbeda dengan Naive Bayes yang memiliki akurasi yang lebih stabil pada kedua data sampel yang besar ataupun kecil.

#### DAFTAR PUSTAKA

- [1] A. Rasool, R. Tao, K. Marjan, and T. Naveed, “Twitter Sentiment Analysis: A Case Study for Apparel Brands Twitter Sentiment Analysis : A Case Study for Apparel Brands,” *IOP Conf. Ser. J. Phys. Conf. Ser.*, 2019, doi: 10.1088/1742-6596/1176/2/022015.
- [2] A. M. Khattak *et al.*, “Tweets classification and sentiment analysis for personalized tweets recommendation,” *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/8892552.
- [3] C. Bhagat and D. Mane, “Survey On Text Categorization Using Sentiment Analysis,” vol. 8, no. 08, 2019.
- [4] I. M. Noor and M. Turan, “Sentiment Analysis using Twitter Dataset,” *IJID (International J. Informatics Dev.*, vol. 9, no. 2, pp. 84–94, 2020, doi: 10.14421/ijid.2019.08206.
- [5] A. Shelar and C. Y. Huang, “Sentiment analysis of twitter data,” *Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018*, pp. 1301–1302, 2018, doi: 10.1109/CSCI46756.2018.00252.
- [6] S. Gohil, S. Vuik, and A. Darzi, “Sentiment analysis of health care tweets: Review of the methods used,” *JMIR Public Heal. Surveill.*, vol. 20, no. 4, 2018, doi: 10.2196/publichealth.5789.
- [7] M. A. Kausar, A. Soosaimanickam, and M. Nasar, “Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, pp. 415–422, 2021, doi: 10.14569/IJACSA.2021.0120252.
- [8] V. A. and S. S. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques,” *Int. J. Comput. Appl.*, vol. 139, no. 11, pp. 5–15, 2016, doi: 10.5120/ijca2016908625.
- [9] M. Kiruthika, W. Sanjana, and P. Giri, “Sentiment Analysis of Twitter Data,” *Int. J. Innov. Eng. Technol.*, vol. 6, no. 4, pp. 2319–1058, 2016,

- [Online]. Available: <http://ijiet.com/wp-content/uploads/2016/04/37.pdf>.
- [10] Y. N. and M. S, "A Review on Text Mining in Data Mining," *Int. J. Soft Comput.*, vol. 7, no. 2/3, pp. 01–08, 2016, doi: 10.5121/ijsc.2016.7301.
- [11] K. B. Priya Iyer and S. Kumares, "Twitter sentiment analysis on coronavirus outbreak using machine learning algorithms," *Eur. J. Mol. Clin. Med.*, vol. 7, no. 3, pp. 2663–2676, 2020.
- [12] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- [13] E. S. Negara, R. Andryani, and P. H. Saksono, "Analisis Data Twitter: Ekstraksi dan Analisis Data Geospasial," *J. INKOM*, vol. 10, no. 1, p. 27, 2016, doi: 10.14203/j.inkom.433.
- [14] I. Saputra, J. A. Halomoan, A. B. Raharjo, C. Rezky, A. Syavira, and S. Informasi, "SENTIMENT ANALYSIS ON TWITTER OF PSBB EFFECT USING MACHINE," pp. 143–150.
- [15] S. Afrizal, H. N. Irmanda, N. Falih, and I. N. Isnainiyah, "Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit," *J. Inform.*, vol. 4221, pp. 157–168, 2019.
- [16] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [17] N. Novalita, A. Herdiani, I. Lukmana, and D. Puspendari, "Cyberbullying identification on twitter using random forest classifier," *J. Phys. Conf. Ser.*, vol. 1192, no. 1, 2019, doi: 10.1088/1742-6596/1192/1/012029.
- [18] S. Goyal, "Review Paper on Sentiment Analysis of Twitter Data Using Text Mining and Hybrid Classification Approach," *International J. Eng. Dev. Res.*, vol. 5, no. 2, pp. 2321–9939, 2017, [Online]. Available: <http://www.ijedr.org/papers/IJEDR1702032.pdf>.
- [19] R. H. Satrio and M. A. Fauzi, "Klasifikasi Tweets Pada Twitter Menggunakan Metode K-Nearest Neighbour ( K-NN ) Dengan Pembobotan TF-IDF," vol. 3, no. 8, pp. 8293–8300, 2019.
- [20] M. Shoeb and J. Ahmed, "Sentiment Analysis and Classification of Tweets Using Data Mining," *Int. Res. J. Eng. Technol.*, vol. 4, no. 12, pp. 1471–1474, 2017, [Online]. Available: [www.irjet.net](http://www.irjet.net).