

## ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP BRAND INDOSAT MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Lady Agustin Fitriana <sup>1</sup>, Safitri Linawati <sup>2</sup>, Nuraeni Herlinawati <sup>3</sup>,  
Rabiatus Sa'adah <sup>4</sup>, Syarah Seimahuria <sup>5</sup>

<sup>1,2,3,4</sup> Sistem Informasi, Universitas Bina Sarana Informatika

<sup>5</sup> Sains Data, Universitas Nusa Mandiri

Jalan Kramat Raya, Jakarta Pusat, Indonesia

*lady.lag@bsi.ac.id*

### ABSTRAK

Era digital saat ini, media sosial memainkan peran penting sebagai platform untuk berekspresi, berbagi informasi dan berkomunikasi dengan orang lain. Twitter merupakan salah satu sosial media yang populer digunakan oleh masyarakat untuk mengeluarkan keluhan. Kelebihan Twitter memungkinkan respons yang cepat dari pihak yang diadukan. Indosat, sebagai perusahaan penyedia layanan telekomunikasi dan jaringan, menawarkan opsi prabayar dan pascabayar dengan memiliki hak dagang IM3 Ooredoo. Tujuan penelitian adalah mengidentifikasi persepsi masyarakat terhadap brand Indosat berupa komentar positif, maupun negatif dengan menggunakan metode *Naïve Bayes Classifier (NBC)* sehingga membantu perusahaan mengetahui karakteristik pelanggan, persepsi kualitas operator Indosat, serta tingkat kepuasan pelanggan untuk meningkatkan strategi marketing yang lebih baik. Penelitian ini menggunakan Algoritma *Naïve Bayes Classifier (NBC)* sebagai metode klasifikasi dan metode CRISP-DM sebagai tahapan penerapan sebuah penelitian. Hasil penelitian menunjukkan bahwa pengguna twitter banyak memberikan "sentimen positif". Hasil pengolahan 100 data tweet's yang terdiri dari 500 *review* positif dan 500 *review* negatif mendapatkan nilai *accuracy* 83,91% +/-3.86% dan *micro average* 83.90 %.

**Kata kunci :** *Sentimen Analisis, Naïve Bayes (NBC), Twitter, Text Mining*

### 1. PENDAHULUAN

Indosat, sebagai perusahaan penyedia layanan telekomunikasi dan jaringan, menawarkan opsi prabayar dan pascabayar dengan memiliki hak dagang IM3 Ooredoo. Perusahaan ini menyajikan layanan MIDI (Multimedia, Internet & Data Communication Services)[1]. Dengan jumlah pelanggan yang signifikan, Indosat berkomitmen untuk memberikan layanan optimal dan promosi menarik, dengan tujuan agar pelanggan dapat merasakan kinerja terbaik saat menggunakan layanannya.

Era digital saat ini, media sosial memainkan peran penting sebagai platform untuk berekspresi, berbagi informasi dan berkomunikasi dengan orang lain. Twitter merupakan salah satu sosial media yang populer digunakan oleh masyarakat untuk mengeluarkan keluhan. Kelebihan Twitter memungkinkan respons yang cepat dari pihak yang diadukan. Dalam beberapa kasus, ketika seseorang mengalami ketidakpuasan atau masalah dengan suatu layanan atau produk, mereka mungkin berharap mendapatkan tanggapan atau solusi secepat mungkin. Twitter memberikan platform yang lebih *real-time* daripada beberapa media sosial lainnya. Tweet yang mengandung keluhan atau emosi kuat dapat dengan cepat menjadi viral atau mendapatkan perhatian banyak orang. Ini dapat menciptakan tekanan tambahan pada perusahaan atau individu yang menjadi subjek keluhan untuk segera merespons atau menanggapi situasi tersebut.

Twitter salah satu sosial media di Indonesia menjadi tumpuan masyarakat dalam menumpahkan opini, tak ayal pengguna aktif twitter pun sangat banyak. Bahkan menurut Kemenkominfo, Indonesia menempati urutan peringkat ke -5 terbesar di dunia sebagai pengguna aktif twitter, bersandingan dengan negara Amerika, Brazil, Jepang dan Inggris[2]. Akun Indosat Ooredoo Care merupakan akun twitter yang memberikan informasi serta menampung dan membalas cuitan komentar dari pelanggannya. Komentar pelanggan dapat berupa kritikan, saran, keluhan bahkan makian dalam mengekspresikan opininya terhadap layanan yang diberikan oleh Indosat.

Analisis dapat menentukan sentiment publik secara manual terhadap layanan operator Indosat. Namun, semakin meningkatnya opini di Twitter dan penataan data, dibutuhkan lebih banyak waktu dan upaya untuk mengklasifikasikan polaritas opini tersebut. Sehingga diperlukan sebuah metode untuk mengekstrak atribut dalam dokumen atau teks yang berisi anotasi untuk menentukan ekspresi dan mengklasifikasikannya sebagai komentar positif atau negatif [3].

*Text Mining* adalah bagian dari penambangan data yang prosesnya adalah mengekstraksi pengetahuan dan informasi dari pola-pola yang ada dalam kumpulan dokumen teks dengan menggunakan alat analisis tertentu. *Text mining* juga dapat digunakan untuk beberapa jenis kebutuhan, antara lain peringkasan, pencarian dokumen teks, dan analisis sentimen[4]. Klasifikasi sentimen adalah

proses mengidentifikasi dan mengelompokkan sentimen atau opini dalam teks ke dalam kategori tertentu, seperti positif, negatif, atau netral. Dalam konteks analisis sentimen, klasifikasi sentimen bertujuan untuk memahami sikap atau perasaan yang terkandung dalam suatu teks, seperti ulasan produk, komentar di media sosial, atau artikel berita. Pendekatan ini melibatkan penggunaan algoritma mesin pembelajaran dan teknik pengolahan bahasa alami untuk secara otomatis menentukan kategori sentimen yang tepat berdasarkan konteks dan kata-kata yang digunakan dalam teks tersebut[5].

Analisis sentimen adalah cara mengumpulkan opini masyarakat dengan menggunakan jejaring sosial yang berisi layanan publik dan topik yang sedang tren. Analisis sentimen mengekstraksi opini publik terhadap topik, produk, atau layanan tertentu yang mengandung teks tidak terstruktur. Tujuan dari analisis sentimen adalah untuk mengetahui apakah pendapat seseorang dalam arti positif atau negatif, yang kemudian dapat digunakan untuk mengambil keputusan[6].

Metode *Naïve Bayes Classifier (NBC)* merupakan metode statistik sederhana dengan akurasi yang baik dan tingkat kesalahan seminim mungkin. *Naïve Bayes Classifier (NBC)* juga dikenal sebagai multinomial *naïve bayes*, yang artinya metode klasifikasi menggunakan metode probabilitas dan statistik. Metode ini merupakan model sederhana dari algoritma Bayesian, cocok untuk klasifikasi teks atau dokumen. Dalam pengklasifikasian menggunakan *naïve bayes*, nilai kategori dokumen akan ditentukan berdasarkan ciri/kata yang muncul pada dokumen yang akan diklasifikasi[7].

Banyak penelitian yang telah menerapkan algoritma *Naïve Bayes Classifier (NBC)* pada *Text Mining* seperti *Sentiment Analysis* atau pengkategorian. Pada penelitian yang berjudul Analisis Sentimen Terhadap Operator Seluler Telkomsel Menggunakan Algoritme *Naïve Bayes* peneliti melakukan klasifikasi sentiment untuk mencari *preference value*. Hasil penelitian menggunakan *cross validation* dan pengukuran akurasi menggunakan *confesion matrix* dengan dilakukan 10 kali pengujian menghasilkan akurasi 85.33% dan respon positif yang diperoleh dari hasil perhitungan *preference values* sebesar 37.03%[8].

Penelitian selanjutnya yang dilakukan oleh Dedi Darwis tentang penerapan Algoritma *Naïve Bayes* untuk sentiment analisis review data twitter pada BMKG untuk mengklasifikasikan data dari Twitter ke dalam opini positif, netral, atau negatif.. Proses melibatkan text mining dan klasifikasi tweet ke dalam tiga kelas utama: positif, negatif, dan netral. Klasifikasi ini mempermudah pengguna dalam melihat berbagai opini yang ada. Hasil uji akurasi metode *Naïve Bayes* menunjukkan efektivitas dalam pengklasifikasian sentiment [9]. Penelitian serupa juga diungkapkan oleh Cholid dan Debby dengan menerapkan metode *Naïve Bayes Classifier (NBC)*

dan *Support Vector Machine* untuk analisis sentiment terhadap dampak virus corona di Twitter. Penelitian ini bertujuan untuk mengetahui sentiment yang bernilai positif, netral maupun negative. Hasil penelitian mengumumkan bahwa metode *Naïve Bayes Classifier (NBC)* mampu mengklasifikasikan sentiment dengan nilai akurasi sebesar 81.07%, sedangkan pengujian dengan *Support Vector Machine (SVM)* memberikan nilai akurasi sebesar 79.96%[10].

Berdasarkan penelitian sebelumnya maka penelitian ini untuk mengetahui persepsi masyarakat terhadap operator Indosat. Pada penelitian ini bertujuan untuk mengidentifikasi persepsi masyarakat terhadap brand Indosat berupa komentar positif, maupun negatif dengan menggunakan metode *Naïve Bayes Classifier* sehingga menghasilkan opini yang bersifat positif dan negatif dengan memanfaatkan sosial media twitter. Twitter menjadi media sosial yang memungkinkan penggunanya mengungkapkan pendapat dan perasaannya terhadap berbagai topik maupun isu. Dengan mengetahui opini atau sentimen oleh pelanggan, dapat membuat perusahaan mengetahui karakteristik pelanggan, persepsi kualitas operator Indosat, serta tingkat kepuasan pelanggan untuk meningkatkan strategi marketing yang lebih baik.

## 2. TINJAUAN PUSTAKA

### 2.1. Text mining

Text Mining adalah proses mendapatkan pengetahuan dan informasi dari pola-pola yang terdapat dalam kumpulan dokumen teks menggunakan alat analisis khusus. Proses ini dapat dimanfaatkan untuk berbagai keperluan, seperti peringkasan, pencarian dokumen teks, dan analisis sentimen[4]. Klasifikasi sentimen adalah proses untuk mengenali dan mengelompokkan sentimen atau opini dalam teks ke dalam kategori seperti positif, negatif, atau netral. Metode ini melibatkan penggunaan algoritma pembelajaran mesin dan teknik pengolahan bahasa alami untuk secara otomatis menentukan kategori sentimen berdasarkan konteks dan kata-kata yang terdapat dalam teks tersebut [11].

### 2.2. Sentimen Analisis

Menurut Bing Liu sentimen analisis melibatkan proses penemuan pola dalam teks yang menunjukkan opini atau sentimen tertentu. Teknik ini menggunakan berbagai metode statistik, linguistik, dan pembelajaran mesin untuk mengkategorikan teks berdasarkan sentimen yang terkandung di dalamnya, seperti positif, negatif, atau netral. Salah satu metode yang sering digunakan dalam sentimen analisis adalah *Naïve Bayes Classifier (NBC)*. Algoritma ini bekerja dengan menghitung probabilitas bahwa sebuah teks termasuk dalam kategori sentimen tertentu berdasarkan fitur-fitur yang ada dalam teks tersebut. NBC adalah pilihan populer karena kesederhanaannya, efisiensi komputasional, dan

performa yang baik dalam berbagai aplikasi sentimen analisis [12].

**2.3. Algoritma Naive Bayes Classifier (NBC)**

Naive Bayes adalah salah satu algoritma yang sering digunakan dalam text mining. Algoritma ini memprediksi kejadian masa depan berdasarkan data dari kejadian sebelumnya, yang dikenal dengan nama Teorema Bayes. Algoritma Naive Bayes didasarkan pada Teorema Bayes, yang menghitung probabilitas suatu kejadian berdasarkan pengetahuan sebelumnya tentang kondisi terkait. Dalam text mining, Naive Bayes digunakan untuk mengklasifikasikan teks ke dalam kategori tertentu (misalnya, positif, negatif, atau netral). Proses klasifikasinya melibatkan langkah-langkah berikut [13]:

- Preprocessing: Membersihkan dan menyiapkan teks, termasuk tokenisasi, penghapusan stop words, dan stemming.
- Pembobotan Fitur: Menghitung frekuensi kemunculan kata (fitur) dalam teks.
- Pembentukan Model: Menggunakan data latih untuk menghitung probabilitas masing-masing kategori berdasarkan fitur-fitur yang ada.
- Klasifikasi: Menggunakan model yang telah dibentuk untuk memprediksi kategori teks baru berdasarkan fitur-fiturnya.

Dasar dari Naive Bayes yang dipakai dalam pemrograman adalah rumus Bayes, yaitu peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B seperti persamaan berikut.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Pada aplikasinya rumus ini berubah menjadi:

$$P(C_i|D) = \frac{P(D|C_i) \times P(C_i)}{P(D)}$$

**2.4. Twitter**

Twitter adalah media sosial sederhana yang memungkinkan pengguna untuk mengunggah pesan singkat atau twit dalam berbagai bentuk, seperti teks, foto, audio, video, dan sebagainya. Twitter merupakan aplikasi yang cukup banyak digunakan oleh masyarakat Indonesia, khususnya dalam memberikan pendapat. Platform ini memfasilitasi komunikasi yang cepat dan mudah, memungkinkan pengguna untuk berbagi informasi, mengemukakan pandangan, serta berinteraksi dengan pengguna lainnya secara langsung. Karena kesederhanaannya dan jangkauan luasnya, Twitter menjadi alat yang efektif untuk menyampaikan kritik, saran, serta opini secara real-time [14].

**2.5. CRISP-DM**

Cross Industry Standard Process for Data Mining (CRISP-DM) adalah model proses independen industri yang sering digunakan untuk keperluan data mining. Metode ini terdiri dari enam fase berulang, yaitu *Business Understanding*, *Data*

*Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [15].

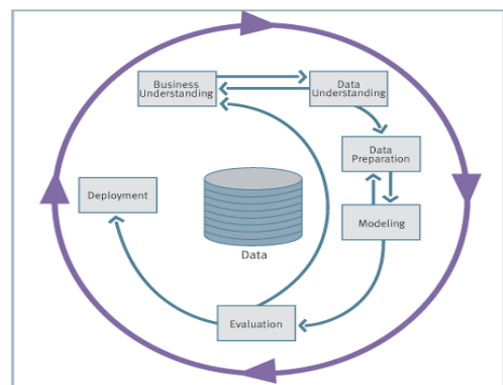
Dalam siklus pengembangannya, CRISP-DM dianggap sebagai metodologi data mining yang paling lengkap untuk memenuhi kebutuhan proyek industri, dan telah menjadi yang paling luas penggunaannya dalam proyek analitik, data mining, serta *data science*. Tahapan penelitian menggunakan CRISP-DM mencerminkan aliran proses dalam model ini, yang umum digunakan untuk data mining. Setiap fase memiliki tugas dan fungsi masing-masing. Panah pada diagram menunjukkan keterkaitan antar fase, namun urutannya bisa fleksibel dan bergantung pada hasil di setiap fase. Lingkaran luar dari setiap fase melambangkan siklus data mining itu sendiri. Model CRISP-DM juga bertujuan untuk membuat proyek penambangan data menjadi lebih murah, lebih andal, lebih dapat diulang, lebih mudah dikelola, dan lebih cepat [16].

Penelitian ini juga menggunakan metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam tahap, yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment. Metodologi ini sering digunakan pada text mining dan memiliki tahapan yang jelas untuk penerapan penelitian data mining [17].

**3. METODE PENELITIAN**

Objek pada penelitian ini merupakan operator Indosat yang saat ini banyak memiliki pelanggan di seluruh Indonesia. Data dari penelitian ini merupakan data kualitatif deskriptif, yakni penelitian berdasarkan pengumpulan fakta atau data berdasarkan kalimat atau tutur kata yang mengandung tweet @indosat. Tweet @indosat inilah yang dikumpulkan sebagai sumber data langsung yang terdapat pada twitter berupa komentar positif ataupun negatif.

Data yang didapatkan dari tweet pengguna Twitter melalui proses crawling dari Twitter API. Dataset berjumlah 1000 data yang terdiri dari label positif dan negatif. Penelitian ini menggunakan Algoritma *Naive Bayes Classifier (NBC)* sebagai metode klasifikasi dan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai tahapan penerapan sebuah penelitian.



Gambar 1. Model CRISP-DM

Metode CRISP-DM pada umumnya sering digunakan pada text mining, metode ini memiliki 6 tahapan, yakni [17]:

a. *Business Understanding*

Tahap ini memahami bagaimana mendapatkan dan membangun data agar mencapai tujuan bisnis. Perspektif bisnis yang dibangun kemudian diubah menjadi pengetahuan *data mining* dengan mengumpulkan data penelitian sebagai langkah awal, kemudian menentukan kapan waktu penelitian yang tepat, serta membuktikan hasil yang diambil sesuai dengan tujuan penelitian.

b. *Data Understanding*

Tahapan ini merupakan proses mengidentifikasi atau memahami data yang akan diuji dalam penelitian. Dalam penelitian ini, data yang diperoleh merupakan sekumpulan data twitt dan komentar dari pengguna akun twitter dengan kata kunci #Indosat.

c. *Data Preparation*

Setelah data melewati tahap data understanding, selanjutnya adakn dilakukan proses preparasi yang terdiri dari beberapa langkah, diantaranya adalah *tokenize*, *filter tokens (by length)*, *stopword removal*, *transform case*, dan *stemming*.

d. *Modelling*

Tahapan ini adalah menentukan algoritma mana yang akan diterapkan dalam melakukan analisis data.

e. *Evaluation*

Tahap evaluasi merupakan penilaian apakah model yang sudah dibuat berdasarkan algoritma yang telah ditentukan menghasilkan sesuai harapan.

f. *Deployment*

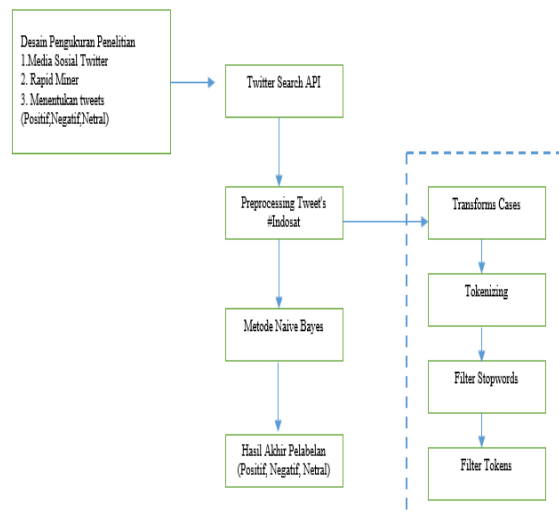
Tahap ini adalah tahap pembuatan atau pembangunan aplikasi sesuai dengan model dan data yang dihasilkan tahapan sebelumnya. Sehingga tahap ini merupakan tahapan akhir dari metode CRISP-DM.

4. HASIL DAN PEMBAHASAN

Penelitian yang dilakukan pada sentiment analisis menggunakan Algoritma *Naïve Bayes Classifier (NBC)* menggunakan aplikasi Rapidminer Studio versi 9.6.

Dalam penelitian ini hal pertama yang penulis lakukan adalah menentukan topik yang dicari dalam media Twitter dengan kata kunci 'indosat', kemudian menghubungkan API Twitter agar mendapatkan akses token untuk *crawling data* dalam *RapidMiner*. Melakukan pelabelan sentiment dengan label *Positive* atau *Negative* sebagai tolak ukur perbandingan penilaian. Perancangan untuk pemilihan text preprocessing dengan *Transforms Case*, *Tokenize*, *Filter Stopwords*, dan *Filter Tokens*. Kemudian dilanjutkan dengan pengujian sentiment menggunakan Algoritma *Naive Bayes Classifier*

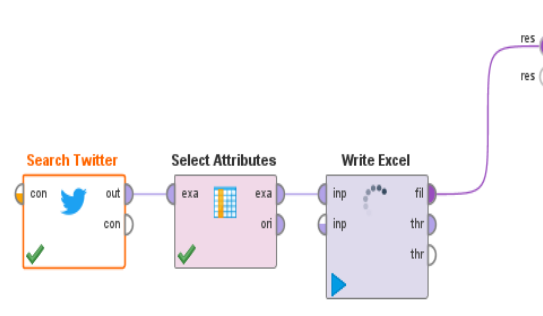
untuk mendapatkan hasil sentiment mesin dengan pelabelan *positive* dan *negative*.



Gambar 2. Kerangka Penelitian

4.1. Pengambilan Data

Pengambilan data penelitian ini menggunakan rapid miner yang sangat membantu peneliti untuk mendapatkan dari twitter.



Gambar 3. Proses Crawling Data

Alur proses *crawling data* atau permulaan mencari data memiliki tiga tahapan proses, yakni proses pertama adalah search twitter dimana pencarian tweet's atau cuitan yang mengandung kata 'indosat' di Twitter untuk dijadikan *dataset sentiment* analisis. Proses kedua merupakan *select attributes* yang merupakan pemilihan dan menghapus atribut yang tidak diperlukan. Sedangkan proses ketiga *write excel* dimana dataset yang telah dikumpulkan diubah menjadi file microsoft excel.

Tabel 1. Alur Proses *Crawling Data*

No.	Proses	Keterangan
1	<i>Search Twitter</i>	Pencarian <i>tweet's</i> kata 'indosat' di Twitter
2	<i>Select Attributes</i>	Pemilihan dan menghapus atribut
3	<i>Write Excel</i>	Dataset diubah menjadi file <i>Ms.Excel</i>

#### 4.2. Data Preparation

Setelah pengumpulan data, data yang telah didapatkan pada twitter tidak dapat diolah secara langsung karena memiliki karakteristik yang tidak terstruktur dan masih banyak memuat *noise*, maka diperlukan *text mining* untuk mengolahnya. Teks mining merupakan proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah data tak terstruktur yang nantinya akan diperoleh pola-pola data, tren dan ekstraksi pengetahuan yang potensial dari data teks. Sehingga dilanjutkan tahapan *data preparation* yang dibagi menjadi beberapa langkah:

##### 4.2.1. Anotation Removal

Langkah ini berguna untuk menghilangkan tanda *anotation* yang terdapat dalam *text*. Perbedaan proses sebelum dan sesudah *anotation removal* dapat dilihat pada tabel berikut:

Tabel 2. Hasil Anotation Removal

Teks	Anotation Removal
indosat jelek banget buat main twitter ganti smartfren lagi ajalah war kalian @smartfrenworld @IndosatCare	indosat jelek banget buat main twitter ganti smartfren lagi ajalah war kalian

##### 4.2.2. Removal URL

*Removal URL* merupakan langkah untuk menghilangkan URL yang terdapat pada *text*.

Tabel 3. Hasil Removal URL

Teks	Removal URL
assalamu'alaikum warahmatullaahi wabarakatuh... min @IndosatCare kok sintalnya dr pagi H+ terus yah? udh restart tp masih gitu saya sudah cek melalui juga <a href="https://bit.ly/twim3wa">https://bit.ly/twim3wa</a>	assalamu'alaikum warahmatullaahi wabarakatuh... min @indosatcare kok sintalnya dr pagi h+ terus yah? udh restart tp masih gitu saya sudah cek melalui juga

##### 4.2.3. Tokenize

Langkah ini merupakan pemisahan kata-kata dari setiap kalimat yang akan diproses. Sehingga kata-kata yang telah diproses dapat dilanjutkan ke langkah berikutnya.

Tabel 4. Hasil Tokenize

Teks	Tokenize
Yg ter--menyebalkan dari susahnya sinyal indosat, hanya chat wa yg bisa dipake jadi emosi doang w gabisa ngerjain yg lain. Migren huy!!	Yg ter--menyebalkan dari susahnya sinyal indosat hanya chat wa yg bisa dipake jadi emosi doang w gabisa ngerjain yg lain Migren huy

##### 4.2.4. Transformation Not (Negative)

Tahapan ini akan menggabungkan dua kata yang berisi kata negatif. Dengan asumsi bahwa kata tersebut tidak ada, jika setelah di *tokenize* kata

tersebut akan dipisahkan dan memiliki makna sendiri. Sehingga untuk menghindari hal itu, dilakukan proses *transformation not (negative)*.

Tabel 5. Hasil Transformation Not (Negative)

Teks	Transformation Not (Negative)
Loh gmn se, yg sy laporkan ini keseluruhan ya. Bukan krn nomor sy saja krn semua keluarga sy pakai indosat. Mohon perhatiannya ya, krn 3 tahun ini sinyal disini benar- benar buruk.	loh gmn se, yg sy laporkan ini keseluruhan ya. bukan_krn nomor sy saja krn semua keluarga sy pakai indosat. mohon perhatiannya ya, krn 3 tahun ini sinyal disini benar- benar buruk.

##### 4.2.5. Stemming

Proses *stemming* merupakan perubahan seluruh kata yang terdapat dalam dataset menjadi kata dasar. Sehingga dapat dilanjutkan dengan proses algoritma klasifikasi.

Tabel 6. Hasil Stemming

Teks	Stemming
Dear @IndosatCare ini kenapa pulsa reguler tetiba kepotong sekian belas ribu? Padahal tidak digunakan telepon/sms, langganan RBT, dan paket internetpun masih ada.	dear @indosatcare ini kenapa pulsa reguler tiba potong kian belas ribu? padahal tidak guna telepon/sms, langgan rbt, dan paket internet masih ada.

##### 4.2.6. Stop Word Removal

Tahapan ini adalah tahap akhir dari proses *data preparation*. Proses ini merupakan penghapusan kata-kata yang tidak memiliki makna atau biasanya berguna sebagai kata sambung ataupun kata keterangan.

Tabel 7. Hasil Stop Word Removal

Teks	Stop Word Removal
dear @indosatcare ini kenapa pulsa reguler tiba potong kian belas ribu? padahal tidak guna telepon/sms, langgan rbt, dan paket internet masih ada.	@indosatcare pulsa reguler potong kian belas ribu? telepon/sms, langgan rbt, paket internet ada.

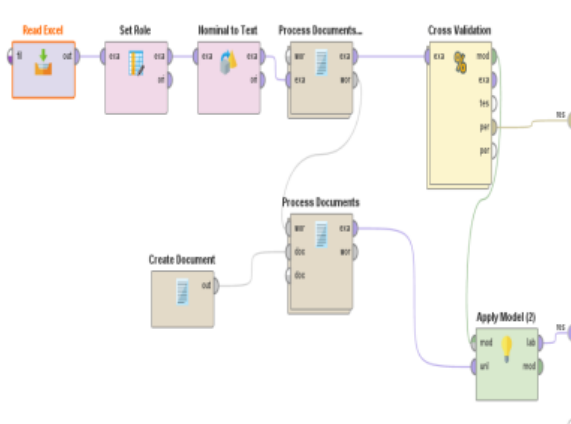
#### 4.2. Modelling

Penelitian ini menggunakan software *rapidminer*, sehingga pemodelan dari keseluruhan penggunaan operator dapat dilihat pada gambar berikut:

Berdasarkan hasil eksperimen yang telah dilakukan pada gambar 4, dapat disimpulkan beberapa hal sebagai berikut:

##### 1. Pelabelan Sentimen

Penulis menggunakan dua pelabelan, label *positive* mengandung unsur kata kepuasan/kebanggaan sedangkan label *negative* mengandung unsur kata kekecewaan ataupun makian.



Gambar 4. Proses Analisis Sentiment

Tabel 8. Data Classification Positive

No	Tweet	Label
1	@IndosatOoredoo Indosat emang hemat prii...sinyal juga lumayan	Positive
2	yaudah, kalo gitu better indosat. kalo xl kuotanya suka bohong ??	Positive

Tabel 9. Data Classification Negative

No	Tweet	Label
1	Indosat kenapa nihhhh, tidak ada sinyal @IndosatCare	Negative
2	indosat jelek banget buat main twitter ganti smartfren lagi ajalah war kalian @smartfrenworld @IndosatCare	Negative

2. Filtering

Hasil dari penelitian menggunakan *Text Processing* yang terdiri dari *Tokenize*, *Transform Cases*, dan *Stopwords*.

Tabel 10. Hasil dari text processing

Text Review	Tokenize	Transform Cases	Stopwords
Yg ter--menyebalkan dari susahny sinyal indosat, hanya chat wa yg bisa dipake jadi emosi doang w gabisa ngerjain yg lain. Migren huy!!	Yg ter--menyebalkan dari susahny sinyal indosat hanya chat wa yg bisa dipake jadi emosi doang w gabisa ngerjain yg lain Migren huy	Yang paling menyebalkan dari susahny sinyal indosat hanya bisa chat sehingga membuat emosi tidak bisa mengerjakan yang lain	sinyal indosat hanya bisa chat

Pada tabel diatas, *Text processing* dari *text review* yang berlabel negatif akan berubah menjadi bersifat positif. Pada *text review* melalui *tokenize* keseluruhan tanda baca seperti koma, titik, tanda seru akan dihilangkan, kemudian pada *transform cases* seluruh *text review* yang menggunakan huruf besar

akan diubah menjadi huruf kecil dan pada *stopwords* kata yang bernilai negatif akan dihilangkan.

Pengujian penelitian menggunakan algoritma *Naive Bayes Classifier (NBC)* mendapatkan hasil sebagai berikut:

Tabel 10. Hasil akurasi

	True Positif	true negative	class precision
pred.positive	898	68	92,96%
pred.negative	134	155	59,63%
class recall	87,02%	69,51%	

5. KESIMPULAN DAN SARAN

Pengujian menggunakan Algoritma *Naive Bayes Classifier (NBC)* dalam sentiment analisis persepsi masyarakat terhadap brand indosat di Twitter menunjukkan bahwa pengguna twitter banyak memberikan “sentimen positif”. Hasil pengolahan 100 data tweet’s yang terdiri dari 500 review positif dan 500 review negatif mendapatkan nilai *accuracy* 83,91% +/-3.86% dan *micro average* 83.90 %. Pada penelitian selanjutnya peneliti dapat mencoba menggunakan beberap algoritma klasifikasi untuk membandingkan berapa akurasi dan keakuratan dalam dataset yang sama dengan menerapkan algoritma *Neural Network*, *Naive Bayes*, *Decision Tree*, *K-NN* dan *Logistic Regression*.

DAFTAR PUSTAKA

- [1] “PT Indosat Tbk.” <https://id.wikipedia.org/wiki/Indosat>.
- [2] Kominfo, “Pengguna Internet di Indonesia 63 Juta Orang,” *Kominfo*. [https://www.kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+internet+di+Indonesia+63+Juta+Orang/0/berita\\_satker](https://www.kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+internet+di+Indonesia+63+Juta+Orang/0/berita_satker).
- [3] A. Kumar dan A. Jaiswal, “Systematic literature review of sentiment analysis on Twitter using soft computing techniques,” *Concurr. Comput. Pract. Exp.*, vol. 32, no. 1, hal. 1–29, 2020, doi: 10.1002/cpe.5107.
- [4] A. Muzaki dan A. Witanti, “Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier,” *J. Tek. Inform.*, vol. 2, no. 2, hal. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [5] I. E. Syahputra, T. Tulus, dan S. Efendi, “Indonesian Text Dataset for Determining Sentiment Classification Using Mechine Learning Approach,” *J. Informatics Telecommun. Eng.*, vol. 3, no. 2, hal. 192–201, 2020, doi: 10.31289/jite.v3i2.3153.
- [6] A. Dwiki et al., “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN,” vol. 8, no. 2, hal. 636–646, 2021.
- [7] A. D. Wibisono, S. Dadi Rizkiono, dan A. Wantoro, “Filtering Spam Email Menggunakan

- Metode Naive Bayes,” *TELEFORTECH J. Telemat. Inf. Technol.*, vol. 1, no. 1, 2020, doi: 10.33365/tft.v1i1.685.
- [8] C. Naya dan Eriswandi, “ANALISIS SENTIMEN TERHADAP OPERATOR SELULER TELKOMSEL MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *Sigma J. Teknol. Pelita Bangsa*, vol. 11, no. 2, hal. 75–82, 2020.
- [9] Dedi Darwis, Nery Siskawati, dan Zaenal Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” *J. TEKNO KOMPAK*, vol. 15, no. 1, hal. 131–145, 2020.
- [10] C. F. Hasri dan D. Alita, “Penerapan Metode Naive Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, hal. 145–160, 2022, doi: 10.33365/jatika.v3i2.2026.
- [11] F. Fathonah dan A. Herliana, “Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes,” *J. Sains dan Inform.*, vol. 7, no. 2, hal. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.
- [12] N. S. Marga, “Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 4, hal. 453–463, 2022, doi: 10.33365/jatika.v2i4.1602.
- [13] D. Duei Putri, G. F. Nama, dan W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, hal. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- [14] Alfandi Safira dan F. N. Hasan, “Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier,” *Zo. J. Sist. Inf.*, vol. 5, no. 1, hal. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
- [15] D. Kurniawan dan M. Yasir, “Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media,” *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 6, no. 2, hal. 74, 2022, doi: 10.22373/cj.v6i2.12793.
- [16] R. KURNIAWAN dan A. APRILIANI, “Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper,” *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 5, no. 1, hal. 67, 2020, doi: 10.24252/instek.v5i1.13686.
- [17] A. Azevedo dan M. F. Santos, “KDD , SEMMA AND CRISP-DM : A PARALLEL OVERVIEW Ana Azevedo and M . F . Santos,” *IADIS Eur. Conf. Data Min.*, hal. 182–185, 2019, [Daring]. Tersedia pada: <http://recipp.ipp.pt/handle/10400.22/136%0Ahttp://recipp.ipp.pt/bitstream/10400.22/136/3/KDD-CRISP-SEMMA.pdf>.