

Analisis Sentimen terhadap Layanan Indihome di Twitter dengan Metode Machine Learning

Rani Puspita^{1*}, Agus Widodo²

^{1,2}Computer Science Department, BINUS Graduate Program – Master of Computer Science,
Bina Nusantara University, Jakarta, Indonesia 11480,
e-mail: ¹rani.puspita@binus.ac.id, ²ag45wd@binus.ac.id

Submitted Date: September 27th, 2021
Revised Date: January 18th, 2022

Reviewed Date: January 05th, 2022
Accepted Date: January 31st, 2022

Abstract

Indihome is a digital service such as the internet that can be used at home, landlines and interactive TV. However, because it is so extensive, Indihome has received a lot of criticism because the internet connection is rarely stable. Therefore, a sentiment analysis in the field of was carried out data mining on customers Indihome on Twitter social media which consisted of 1350 data and filtered into 1309 data because a lot of data indicated duplicates. In this study, researchers used the methods Random Forest and Gradient Boosted Trees (GBT). This research was conducted using tools Rapidminer version 9.8. Research shows that sentiment analysis on Indihome services using the method Random Forest achieves an accuracy of 99.54% with class precision for pred. negative is 99.92%, pred positive is 25.00%, and pred. neutral is 60.00%. Then the GBT method has an accuracy rate of 99.31% with a precision class ofn for pred. negative is 99.46%, pred. positive is 0.00%, and pred. neutral is 0.00%. So it can be concluded that the Random Forest method is a better method when compared to GBT.

Keywords: Sentiment Analysis; Indihome; Twitter; Random Forest; Gradient Boosted Trees

Abstrak

Indihome adalah layanan *digital* seperti internet yang dapat digunakan di rumah, telepon rumah maupun TV Interaktif. Namun karena sudah sangat luas, maka Indihome sangat banyak mendapatkan kritikan karena koneksi internet yang jarang stabil. Maka dari itu, dilakukanlah analisis sentimen di bidang *data mining* terhadap *customer* Indihome pada media sosial Twitter yang terdiri dari 1350 data dan dilakukan *filtering* menjadi 1309 data sebab banyak data yang terindikasi duplikat. Pada penelitian ini, peneliti menggunakan metode *Random Forest* dan *Gradient Boosted Trees* (GBT). Penelitian ini dilakukan dengan *tools* Rapidminer versi 9.8. Penelitian menunjukkan hasil bahwa analisis sentimen terhadap layanan Indihome dengan menggunakan metode *Random Forest* mencapai akurasi sebesar 99.54% dengan *class precision* untuk *pred.* negatif adalah 99.92%, *pred* positif adalah 25.00%, dan *pred.* netral adalah 60.00%. Lalu pada metode GBT tingkat akurasinya mencapai 99.31% dengan *class precision* untuk *pred.* negatif adalah 99.46%, *pred.* positif adalah 0.00%, dan *pred.* netral adalah 0.00%. Maka dapat disimpulkan bahwa metode *Random Forest* merupakan metode yang lebih baik jika dibandingkan dengan GBT.

Kata Kunci: Analisis Sentimen; Indihome; Twitter; *Random Forest*; *Gradient Boosted Trees*

1. Pendahuluan

Internet telah menjadi kebutuhan primer untuk seluruh masyarakat dan tidak dapat ditinggalkan (Fitriasti & Priansa, 2021). PT Telkom adalah sebuah perusahaan telekomunikasi yang berada di Indonesia. PT Telkom juga telah mengeluarkan banyak *product* dengan layanan beragam atau *multiservice*. Maksud dari

multiservice di sini adalah karena produk PT Telkom menyediakan berbagai layanan seperti telepon, gambar, dan data dalam satu jaringan atau disebut juga dengan layanan *Triple Play*. Layanan PT Telkom ini diberi *merk* IndiHome (Suarjana & Suprapti, 2018). IndiHome adalah layanan berupa internet yang dapat digunakan di rumah. Selain internet, tersedia juga layanan seperti telepon, dan

UseTV (Syahputra & Sampurno, 2019). Indihome sangat banyak mendapatkan kritikan. Tetapi banyak pula yang memuji layanan ini. Di era yang serba *digital* ini, kebanyakan *customer* mengemukakan pendapatnya di media sosial Twitter. Twitter sudah berkembang dengan sangat pesat. Terdapat lebih dari 313 juta user aktif per bulan dan terdapat sekitar 500 juta *tweet* per hari pada Twitter. Twitter sangat berpengaruh bagi organisasi maupun individu yang aktif di media sosial untuk mengetahui informasi berupa politik maupun ekonomi (Rasenda, Lubis, & Ridwan, 2020).

Berdasarkan studi pustaka yang peneliti baca dan tuliskan, penelitian mengenai analisis sentimen yang dilakukan oleh Rasenda dkk dengan judul “Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter” (Rasenda et al., 2020) menggunakan Rapidminer. Metode yang digunakan adalah KNN. Proses penelitiannya seperti pelabelan, *pre-processing*, tahap *classification of sentiment analysis* dan evaluasi. Lalu akurasi yang didapatkan adalah 70.59%.

Lalu studi literatur berikutnya dengan judul “Analisa Sentimen *Cyberbullying* Di Jejaring Sosial Twitter Dengan Algoritma *Naïve Bayes*” oleh (Maulana, Ernawati, Labu, & Selatan, 2020) dilakukan menggunakan Python. Metode yang digunakan adalah *Naïve Bayes*. Proses penelitiannya adalah akuisisi, pra proses data, klasifikasi dan evaluasi. Lalu akurasi pada metode ini adalah 76 persen. Hasil penelitian tergolong cukup baik sehingga dapat mengklasifikasikan *tweet* yang bersifat positif maupun bersifat negatif. Tetapi pada proses evaluasi, maka penelitian ini dianggap kurang baik sebab masih banyak kesalahan prediksi dimana isi *tweet* yang sebenarnya tidak mengandung unsur *cyber bullying* tetapi justru memiliki label *tweet* yang negatif dan ini dianggap kurang sinkron.

Selain itu studi literatur dengan judul “Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma *Deep Neural Network*” oleh (Hidayat, Hardiansyah, & Affandy, 2021) dilakukan dengan menggunakan Python. Lalu penelitian ini menggunakan algoritma *Deep Neural Network*. Proses yang dilakukan adalah pengumpulan *dataset tweet*, *cleansing*, *case folding*, formalisasi, *stemming* dan tokenisasi. Kemudian akurasi mencapai 88.72%.

Berdasarkan latar belakang dan studi pustaka yang telah dijelaskan, maka dari itu akan digunakan

dua metode *machine learning* dalam penelitian ini untuk memperbandingkan akurasi dalam penelitian ini. Karena KNN, *Naïve Bayes* dan *Deep Neural Network* menghasilkan akurasi yang tidak begitu baik, maka dari itu peneliti akan menggunakan metode *Random Forest* dan *Gradient Boosted Trees* untuk perbandingan hasil akurasi nantinya.

Metode *machine learning* merupakan metode yang digunakan untuk mengekstraksi data atau bahkan menemukan pola dalam data yang ada. *Machine learning* adalah metode melakukan prediksi di masa yang akan datang berdasarkan data yang ada di masa lalu (Putra & Santika, 2020).

Metode pertama yang akan digunakan adalah *Random Forest*. *Random forest* merupakan sebuah metode *ensemble learning* oleh Breiman pada tahun 2001. *Random Forest* merupakan sebuah kombinasi dari klasifikasi sehingga setiap pohon akan bergantung pada nilai acak vektor secara mandiri. *Random Forest* biasa digunakan untuk klasifikasi dan juga regresi. Hal ini sejalan dengan pendapat (Christy & Suryowati, 2021) yang menyatakan bahwa *Random Forest* juga merupakan proses *classification* yang terdiri dari beberapa pohon keputusan. Langkah-langkah metode *Random Forest* yaitu diantaranya: (Syukron & Subekti, 2018):

1. Menghasilkan *training* baru dengan sampel yang acak.
2. Pada setiap *training* yang baru, maka dibangunlah sebuah pohon dengan pemilihan fitur secara acak juga.
3. Setelah menghasilkan sejumlah pohon, maka data segera diprediksi dengan menggabungkan hasil dari semua pohon dengan *voting*.

Lalu selain *Random Forest*, maka metode kedua yang akan digunakan adalah Metode *Gradient Boosted Trees* (GBT). GBT merupakan metode klasifikasi yang menangani data numerik, nominal, maupun *missing data*. GBT biasanya membangun *decision tree* sesuai dengan peningkatan pada struktur pohon di pembelajaran yang lemah untuk memperbaiki kesalahan pohon dan juga mencegah terjadinya *overfitting*. Jika membangun *decision tree*, maka dapat dilakukan penambahan jumlah iterasi sehingga hal tersebut dapat meningkatkan kinerja model. GBT dalam hal ini sangat mampu dalam memecahkan masalah dengan menyesuaikan pembelajaran dengan gradien negatif dari *loss function*, lalu meningkatkan *trees* dengan parameter yang mewakili (Bisri & Rachmatika, 2019).

Oleh karena itu, penelitian dilakukan menggunakan metode dua metode yaitu *Random*

Forest dan GBT menggunakan data yang diambil dari Twitter dan menggunakan *tools* Rapidminer untuk analisis datanya.

Rumusan permasalahan dalam penelitian ini adalah bagaimana caranya melakukan *pre processing* data dan seberapa besar tingkat akurasi dari metode *Random Forest* dan GBT. Selain itu, terdapat tujuan dalam penelitian ini dimana tujuannya adalah untuk mengetahui bagaimana teknik *pre processing* yang baik dan juga untuk mengetahui seberapa besar tingkat akurasi dari kedua metode yang diteliti serta untuk mengetahui metode mana yang akurasinya paling baik.

Penelitian ini merupakan sebuah ide yang peneliti buat sebagai bahan penelitian untuk analisis sentimen terhadap layanan Indihome. Peneliti menggunakan *tools* Rapidminer lalu dilanjutkan dengan menganalisis kedua metode dengan *tools* yang telah ditentukan.

2. Metodologi Penelitian

2.1 Metode Pengumpulan Data

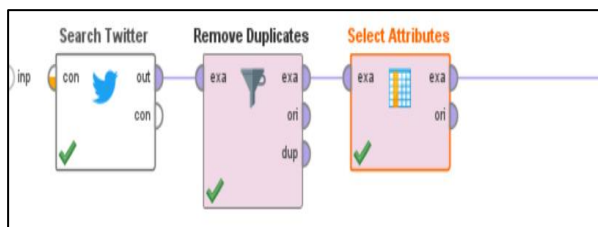
Pengumpulan data dilakukan dengan beberapa cara, yaitu:

a. Studi Pustaka

Peneliti membaca banyak jurnal baik jurnal nasional maupun jurnal internasional. Setelah mendapatkan referensi terhadap banyak jurnal, setelah itu peneliti mencari informasi yang dibutuhkan dalam melakukan penelitian ini. Peneliti juga mencantumkan daftar pustaka yang dapat dilihat pada halaman daftar pustaka.

b. Pengumpulan Data di Media Sosial Twitter

Pada tahap ini, peneliti mengumpulkan data dari media sosial Twitter. *Tools* yang digunakan adalah RapidMiner versi 9.8. Gambar 1 adalah ilustrasi proses *crawling* dan *filtering* data dari Twitter menggunakan RapidMiner.



Gambar 1. *Crawling* Dan *Filtering* Data dari Twitter

Tahap awal yang dilakukan pada pengumpulan data twitter adalah *crawling* data pada media sosial Twitter dengan jumlah data sebanyak 1350. Kemudian proses selanjutnya

adalah *filtering* data yang dilakukan dengan cara menghapus data yang terindikasi duplikat. Setelah melakukan *crawling* dan *filtering*, maka peneliti mendapatkan data akhir sebanyak 1309 data yang kemudian dilanjutkan dengan menyeleksi atribut yang akan digunakan guna memilih *text* pada data. Lalu data di *eksport* dalam bentuk *excel* dan dilakukanlah pelabelan data yang terdiri dari tiga bagian yaitu yang bersifat positif, negatif atau netral.

2.2 Alat yang Digunakan Pada Penelitian

Alat yang digunakan dalam penelitian terbagi menjadi dua, yaitu *hardware* dan *software*. Berikut rincian peralatan penelitian yang digunakan:

a. *Hardware*:

1. HP Laptop 14s-cf0055TX.
2. RAM 8 GB.

b. *Software*:

1. Sistem Operasi Windows 10 Pro 64bit.
2. RapidMiner versi 9.8.
3. Sistem operasi windows 10 profesional.
4. Microsoft Excel.

2.3 *Crawling* Data Twitter

Row No.	Id	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text
1	1440922345...	Sep 23, 2021...	Willy >	147095418	kechwebong	147095418	in	<a href="http://...	Gak inc...
2	1440917569...	Sep 23, 2021...	Sandjaja Mr	25770767	?	-1	in	<a href="http://...	Mama
3	1440917239...	Sep 23, 2021...	Jimbo	9537150155...	AuliaRachma	73824039	in	<a href="http://...	@Aulia
4	1440910577...	Sep 23, 2021...	J Dini Kumal...	225453306	happippe	60243298	in	<a href="http://...	@happ
5	1440907882...	Sep 23, 2021...	Denis	333895290	?	-1	in	<a href="http://...	Kosan
6	1440902767...	Sep 23, 2021...	Wendi Noorc...	1990335308	ridwanhr	43922847	in	<a href="http://...	@ridwa
7	1440902287...	Sep 23, 2021...	EAT	204814909	?	-1	in	<a href="http://...	hissaja
8	1440894764...	Sep 23, 2021...	akun tidak ter...	1259524345...	wkisp	1153551288...	in	<a href="http://...	@wksis
9	1440893002...	Sep 23, 2021...	Komala Som...	1291201863...	?	-1	in	<a href="http://...	indiehc
10	1440892099...	Sep 23, 2021...	Dad on Duty	213470453	tubifess	8857777016...	in	<a href="http://...	@tubif
11	1440891045...	Sep 23, 2021...	sibilusmini	1357954979...	?	-1	in	<a href="http://...	indiehc
12	1440891017...	Sep 23, 2021...	ngashim	69225472	otranel	486218941	in	<a href="http://...	@otrar

Gambar 2. *Crawling* Data Twitter

Pada Gambar 2, terdapat banyak atribut diantaranya adalah atribut *Id*, *Created-At*, *From-User*, *language*, *source*, *logitude*, *latitude*, *text* dan lain sebagainya. Dalam hal ini, peneliti hanya membutuhkan atribut *text*, maka dari itu peneliti me-remove atribut selain *text*.

Row No.	Id	Text
1	1440922345519637959	Gak indihome gak telomsel sama aja sinyalnya ****
2	1440917599087496198	Mama papa lagi keluar abang indihome datang https://t.co/ZoZorUgH7GI
3	1440917239322976257	@hullaRachma @mukareza Aku telomsel dan indihome fine2 saja
4	1440910577862279168	@nappipe Haaap iya sama aku kira indihome tapi ternyata tadi aku hubungin @jeniushelp lagi kendala
5	1440907862640261123	Kosan pale indihome aku pun jadi sial
6	1440902767053197312	@indwantr Suka aneh yg ngeluh indihome kenapa ga ganti provider aja
7	1440902287182893056	hilsajan indihome ki noen masuk kok.
8	1440894764254052352	@vossp bulan beritue mboret soale indihome trobel
9	1440893002830860288	indihome lagi rajin banget nawarin nambahin kecepatan, tapi emang belakangan internet lagi lemot banget. senjaja kali
10	1440892069514997507	@bubifress Padahal videonya bukan lagi nyelam benetin kabel indihome. @gig
11	1440891945248585728	indihome km gacapa apa ya bikin semua orang emosi??
12	1440891017905926146	@citranel @wirabattana @_rozam @JogjalUpdate @IndiHome Ngaplik ga ada nih, padahal udh request di survey katanya.

Gambar 3 Crawling Data Twitter Final

Gambar 3 adalah *screenshot* dari data yang telah berhasil di *crawling* sehingga hanya terdapat beberapa atribut saja seperti *row no*, *id*, dan *text* saja.

2.4 Tahap Pre-processing

Data yang diperoleh dari hasil *crawling* data Twitter adalah sebanyak 1309 *record* mengenai layanan Indihome dengan *tools* RapidMiner. Teknik *pre-processing* yang digunakan adalah sebagai berikut:

a. Data Validasi

Pada tahap ini, peneliti melihat dan *remove* atau menghapuskan data yang tidak digunakan seperti data yang tidak konsisten, dan juga *missing value*.

b. Data Integration dan Transformation

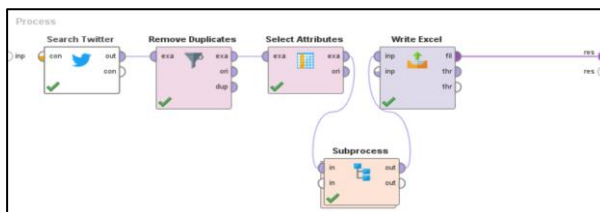
Pada tahap ini, akan dilakukan eksperimen yang bertujuan untuk meningkatkan akurasi dari metode-metode yang digunakan.

c. Data Size Redution dan Dcretization

Pada tahap ini, peneliti mempersiapkan data yang telah di *collect* dengan cara *remove* data yang terindikasi duplikat.

2.4.1 Export Data ke Excel

Pada tahap ini, maka akan dilakukan *export* data menjadi *excel* agar dapat dilakukan analisa pada *tools* RapidMiner. Gambar 4 adalah ilustrasi untuk *export data to excel*.

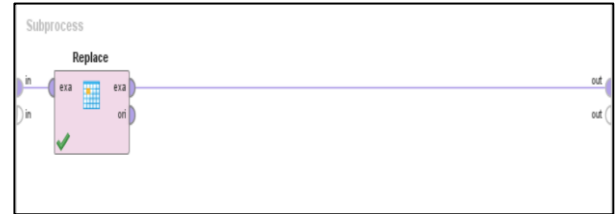


Gambar 4 Export Data ke Excel

Gambar 4 adalah proses yang dilakukan pada RapidMiner untuk meng-*export* data menjadi *excel*. Pada ilustrasi tersebut, maka dapat dilihat bahwa terdapat banyak atribut seperti *Search*

Twitter, *Remove Duplicate*, *Replace Attributes*, *Sub Process*, dan *Write Excel*.

2.4.2 Sub Proses Penelitian



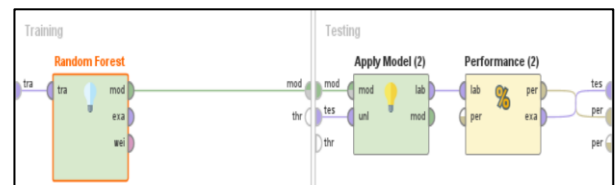
Gambar 5 Sub Proses Penelitian

Gambar 5 adalah contoh *sub process* pada penelitian. Terdapat *action* untuk *me-replace* dimana hal ini bertujuan untuk menghilangkan kata yang tidak akan digunakan seperti, @, dan lain-lain.

2.5 Tahap Cross Validation

Tahap ini adalah sebuah tahapan yang dilakukan untuk mencari akurasi dari masing-masing metode. Pada tahap *cross validation* dalam Rapidminer, maka dibagi menjadi dua *column* yaitu *training* dan *testing*. Untuk *training*, terdapat *action* terdapat metode yang digunakan dan untuk *testing* dapat berupa *apply model* dan *performance*.

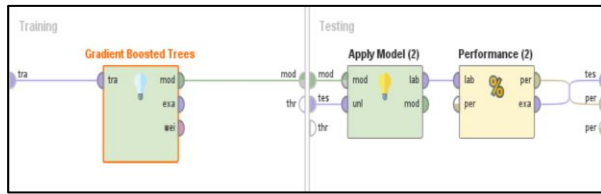
2.5.1 Tahap Cross Validation Metode Random Forest



Gambar 6 Cross Validation dengan Metode Random Forest

Gambar 6 adalah tahap yang dilakukan untuk *cross* dalam metode *Random Forest*. Pada kolom *training* terdapat *action* untuk metode *Random Forest*. Dan pada kolom *testing* terdapat *action* berupa *apply model* dan juga *performance*. Berdasarkan hal tersebut, maka tujuannya adalah untuk mengetahui seberapa besar akurasi pada setiap metode yang sudah di *testing*.

2.5.2 Cross Validation Gradient Boosted Trees (GBT)

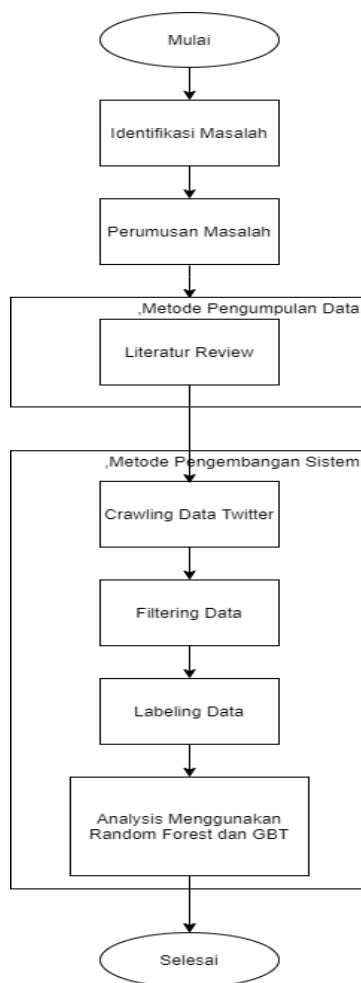


Gambar 7 Cross Validation Metode GBT

Gambar 7 adalah cara yang peneliti lakukan untuk melakukan *cross validation* menggunakan metode GBT. Pada kolom *training* terdapat *action* untuk metode GBT. Dan pada kolom *testing* terdapat *action* berupa *apply model* dan juga *performance*. Berdasarkan hal tersebut, maka tujuannya adalah untuk mengetahui seberapa besar akurasi pada metode GBT.

2.6 Tahapan Penelitian

Gambar 8 merupakan ilustrasi terkait tahapan penelitian:



Gambar 8 Tahapan Penelitian

Berikut ini adalah penjelasan Gambar 8 terkait tahapan penelitian:

a. Mengidentifikasi Permasalahan

Pada *step* ini, maka akan dilakukan identifikasi terhadap masalah yang kemudian akan diselesaikan dalam penelitian ini.

b. Rumusan Permasalahan

Rumusan permasalahan dalam penelitian ini adalah bagaimana cara meng-*collect* data melalui media sosial Twitter menggunakan RapidMiner dan seberapa besar tingkat akurasi metode *Random Forest* dan GBT.

c. Pengumpulan Data

Pada *step* ini, maka dilakukan studi literatur. Studi literatur yang dibaca dan dituliskan dapat berupa jurnal internasional dan juga jurnal nasional.

d. Crawling Data

Pada *step* ini, data yang dikumpulkan sebanyak 1350 data.

e. Data Filtering

Pada *step* ini, akan dilakukan *filtering* sebab terdapat banyak data yang terindikasi duplikat. Maka dari itu, diketahui jumlah data awal adalah 1350, maka setelah dilakukan *filtering*, data yang akan digunakan dalam eksperimen adalah sebanyak 1309 data.

f. Pelabelan Data

Pada *step* ini, akan dilakukan proses pelabelan data agar mengetahui karakteristik dari teks yang sudah di-*collect*. Pelabelan terbagi menjadi tiga bagian yaitu label positif, negatif, dan juga netral.

g. Data Analisis

Pada *step* ini, dilakukan analisis untuk mengetahui tingkat akurasi dari kedua metode yang digunakan yaitu metode *Random Forest* dan GBT.

3. Hasil dan Pembahasan Penelitian

Peneliti melakukan analisis Sentimen terhadap Indihome dengan menggunakan metode *Random Forest* dan GBT dengan menggunakan *tools* RapidMiner. Maka akan ditampilkan beberapa *screenshot* dari *crawling* data yang sudah dilabelkan dan terdapat juga penjelasan untuk nilai akurasi dari setiap metode yang sudah dipilih sebelumnya.

3.1 Labeling

Row No.	Sentiment	Text	Id
1	neu	@Restumamaa bangkeee ngucapinya at twitter wk. maci eka anaq indiehome ses..	1440925701..
2	neu	Wah klo pale indiehome msh berek bun netfanya ??	1440925211..
3	neu	Indiehome kan punya kemampuan menerawang masa depan, kenapa mereka bda..	1440924803..
4	neg	Gak indiehome gak telomsel sama aja siyahnya ??????	1440922345..
5	neu	Mama papa lagi keluar abang indiehome datang https://t.co/ZoRuiqITGI	1440917569..
6	pos	@KuliaRachma @mukareza Aku telomsel dan indiehome fine2 saja	1440917239..
7	pos	@hasipope Hasip ya sama aku kira indiehome tapi ternyata tadi aku hubungin @jen..	1440910577..
8	neg	Kosan pake Indiehome aku pun jadi sial	1440907882..
9	pos	@idwantr Suka aneh yg ngeluh Indiehome kenapa ga ganti provider aja	1440902767..
10	neg	hiisan indiehome hi ncan manuk kok.	1440902287..
11	neg	@wasip bulan bentele mrebet soale indiehome tobel	1440894764..
12	neg	indiehome lagi rajin banget nawarin nambahin kecepatan, tapi emang belakangan l..	1440893002..

Gambar 9 Pelabelan Data

Gambar 9 adalah ilustrasi dari pelabelan data. Diketahui bahwa data sebelum dilabel berupa *row*, *Id*, dan *text*. Oleh karena itu, pada tahap ini ditambahkan satu kolom lagi yang bernama *sentiment* yang bertujuan untuk mengetahui karakteristik dari *text* yang sudah di *crawling*. Ada tiga karakteristik dalam eksperimen yaitu data yang bersifat positif, negatif, dan juga netral.

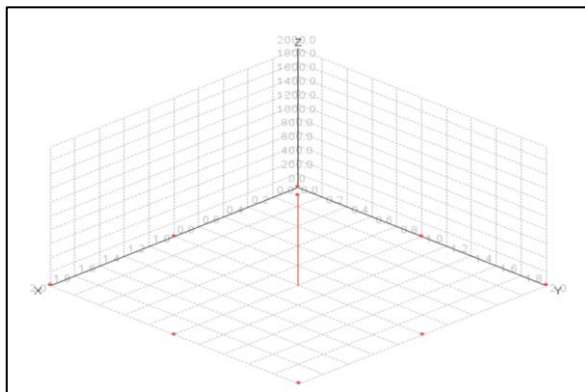
3.2 Akurasi Metode Random Forest

accuracy: 99.54% +/- 0.39% (micro average: 99.54%)				
	true neu	true neg	true pos	class precision
pred. neu	3	1	1	60.00%
pred. neg	0	1299	1	99.92%
pred. pos	1	2	1	25.00%
class recall	75.00%	99.77%	33.33%	

Gambar 10. Hasil Akurasi Metode Random Forest

Hasil dari metode *Random Forest* menghasilkan akurasi sebesar adalah 99.54%. Diketahui bahwa *class precision* untuk *pred* negatif adalah 99.92%, *pred* positif adalah 25.00%, dan *pred*. netral adalah 60.00%.

Gambar 11 adalah ilustrasi terkait *plot view* dari metode ini. Berikut ilustrasi nya:



Gambar 11. Plot View Menggunakan Metode Random Forest

Selain itu, terdapat juga performa vektor untuk metode *Random Forest*.

PerformanceVector			
PerformanceVector:			
accuracy: 99.54% +/- 0.39% (micro average: 99.54%)			
ConfusionMatrix:			
True:	neu	neg	pos
neu:	3	1	1
neg:	0	1299	1
pos:	1	2	1

Gambar 12. Performa Vektor Menggunakan Metode Random Forest

Performa vektor merupakan deskripsi dari tabel hasil analisis yang telah peneliti lakukan.

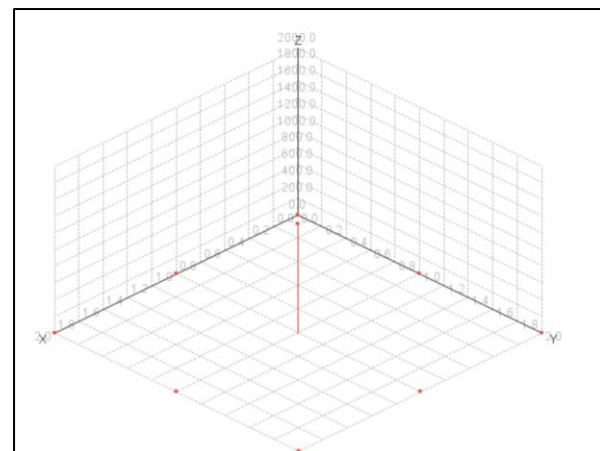
3.3 Akurasi Metode GBT

accuracy: 99.31% +/- 0.43% (micro average: 99.31%)				
	true neu	true neg	true pos	class precision
pred. neu	0	2	0	0.00%
pred. neg	4	1300	3	99.46%
pred. pos	0	0	0	0.00%
class recall	0.00%	99.85%	0.00%	

Gambar 13. Hasil Akurasi Metode GBT

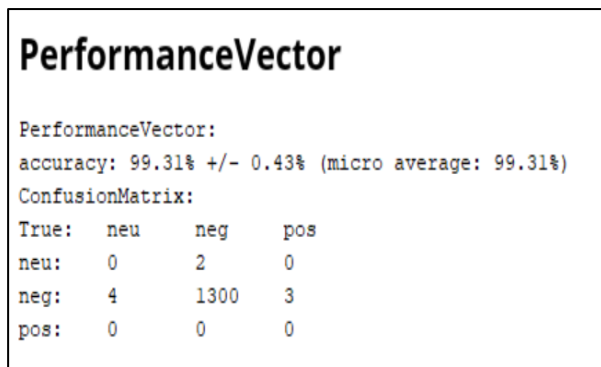
Hasil dari metode GBT menghasilkan akurasi sebesar 99.31%. Diketahui bahwa *class precision* untuk *pred*. negatif adalah 99.46%, *pred*. positif adalah 0.00%, dan *pred*. netral adalah 0.00%.

Lalu terdapat ilustrasi terkait *plot view* dari metode GBT.



Gambar 14. Plot View Menggunakan Metode GBT

Selain itu, terdapat performa vektor pada metode GBT yang dapat dilihat pada Gambar 15.



```
PerformanceVector
PerformanceVector:
accuracy: 99.31% +/- 0.43% (micro average: 99.31%)
ConfusionMatrix:
True:  neu   neg   pos
neu:   0     2     0
neg:   4    1300   3
pos:   0     0     0
```

Gambar 15. Performa Vektor Menggunakan Metode GBT

Performa vektor merupakan deskripsi dari tabel hasil analisis yang telah peneliti lakukan.

4. Kesimpulan

Penelitian ini menggunakan metode *Random Forest* dan GBT. *Tools* yang digunakan untuk eksperimen adalah RapidMiner versi 9.8. Langkah awal yang dilakukan adalah *crawling* data dari media sosial Twitter dengan menggunakan API Twitter, lalu *filtering*, pelabelan, dan juga analisis data untuk mengetahui seberapa besar akurasi dari masing-masing metode. Metode *Random Forest* dan GBT merupakan metode *machine learning*.

Analisis sentimen terhadap layanan Indihome mencapai tingkat akurasi 99.51%. Diketahui bahwa *class precision* untuk *pred. negatif* adalah 99.92%, *pred positif* adalah 25.00%, dan *pred. netral* adalah 60.00% untuk metode *Random Forest*. Lalu tingkat akurasi mencapai 99.31%. Dimana *class precision* untuk *pred. negatif* adalah 99.46%, *pred. positif* adalah 0.00%, dan *pred. netral* adalah 0.00% dalam metode GBT.

Maka disimpulkan bahwa metode yang terbaik adalah metode *Random Forest* sebab *Random Forest* menghasilkan akurasi yang lebih tinggi dibandingkan dengan metode GBT. Tingkat akurasi untuk *Random Forest* mencapai 99.54%.

5. Saran

Dari penelitian yang telah dilakukan, maka masih terdapat banyak hal yang dapat dikembangkan. Saran yang dapat diberikan untuk pengembangan penelitian ini adalah sebagai berikut:

a. Mempertimbangkan jumlah data yang lebih banyak untuk meningkatkan akurasi dari analisis yang telah dilakukan.

b. Dapat menggunakan metode lainnya untuk analisis sentimen.

References

- Bisri, A., & Rachmatika, R. (2019). Integrasi Gradient Boosted Trees dengan SMOTE dan Bagging untuk Deteksi Kelulusan Mahasiswa. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(4), 309. <https://doi.org/10.22146/jnteti.v8i4.529>
- Christy, E., & Suryowati, K. (2021). Analisis Klasifikasi Status Bekerja Penduduk Daerah. 6(1), 69–76. Retrieved from https://ejournal.akprind.ac.id/index.php/STATIS_TIKA/article/view/3519/2577
- Fitriasti, N. A., & Priansa, D. J. (2021). STRATEGI DIRECT MARKETING DALAM RANGKA MERANGSANG MINAT PENGGUNAAN PRODUK INDIHOME. 7(4), 665–671. Retrieved from <https://openlibrarypublications.telkomuniversity.ac.id/index.php/appliedscience/article/view/15264>
- Hidayat, E. Y., Hardiansyah, R. W., & Affandy. (2021). Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network. 02, 108–118. Retrieved from <https://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/1887/pdf>
- Maulana, F. A., Ernawati, I., Labu, P., & Selatan, J. (2020). Analisa sentimen cyberbullying di jejaring sosial twitter dengan algoritma naïve bayes. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 529–538. Retrieved from <https://conference.upnvj.ac.id/index.php/senamika/article/view/619>
- Putra, A. I., & Santika, R. R. (2020). Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 121–130. <https://doi.org/10.29408/edumatic.v4i1.2162>
- Rasenda, R., Lubis, H., & Ridwan, R. (2020). Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter. *Jurnal Media Informatika Budidarma*, 4(2), 369. <https://doi.org/10.30865/mib.v4i2.2051>
- Suarjana, I. K., & Suprapti, N. W. S. (2018). Pengaruh Persepsi Harga, Pengetahuan Produk, Dan Citra Perusahaan Terhadap Niat Beli Layanan Multi Servis Merek Indihome. *E-Jurnal Manajemen Universitas Udayana*, 7(4), 251833. <https://doi.org/10.24843/EJMUNUD.2018.v07.i04.p08>
- Syahputra, R. D., & Sampurno, W. (2019). Analisis Kualitas Pelayanan Nilai Pelanggan Berbasis Digital pada Media Sosial Facebook terhadap

Kepuasan Pelanggan (Studi Pada Produk Indihome Di Plaza Telkom Lembong Kota Bandung Tahun 2019). *EProceedings of Applied Science*, 5(3), 1841. Retrieved from <https://openlibrarypublications.telkomuniversity.ac.id/index.php/appliedscience/article/view/11042/10910>

Syukron, A., & Subekti, A. (2018). Penerapan Metode

Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit. *Jurnal Informatika*, 5(2), 175–185. <https://doi.org/10.31311/ji.v5i2.4158>