

## KLASIFIKASI KOMENTAR INSTAGRAM UNTUK IDENTIFIKASI KELUHAN PELANGGAN JASA PENGIRIMAN BARANG DENGAN TEKNIK SMOTE

NANANG RUHYANA  
DIDI ROSIYADI

Ilmu Komputer STMIK Nusa Mandiri  
Jalan Kramat Raya Nomor 18 Jakarta Pusat  
Email: 14002086@nusamandiri.ac.id  
didi016@lipi.go.id

**Abstrak.** Kepuasan pelanggan adalah salah satu hal yang diharapkan suatu perusahaan ketika produk yang dihasilkan telah dipasarkan, baik berupa barang maupun jasa. Cara pengaduan melalui *customer service* sangat beragam, belakangan ini tak hanya melalui telepon saja pelanggan menyampaikan saran atau keluhannya. Pelanggan bisa menyampaikan saran atau keluhannya melalui surat elektronik atau e-mail maupun forum-forum di dunia maya yang memang dibuat perusahaan penghasil produk untuk menampung beragam keluhan, saran, dan kritik *langsung* dari konsumen terlebih sosial media yang bebas mengutarakan pendapat mereka terhadap jasa pengiriman yang mereka pakai. Instagram merupakan sebuah sosial media yang lebih cenderung kepada gambar dan disisi lain mempunyai teks caption dan komentar, dari permasalahan tersebut diatas mencoba membuat sebuah penelitian untuk keluhan pelanggan pengguna jasa pengiriman barang pada sebuah akun Instagram perusahaan jasa pengiriman. Dari latar belakang permasalahan itulah peneliti mencoba memecahkan masalah untuk pengklasifikasi *text mining* dengan menggunakan metode *Support Vector Machine* (SVM) dan *Naïve Bayes* dan menggunakan teknis SMOTE dengan proses yang biasa dilakukan untuk *text mining* sehingga dapat menghasilkan akurasi 69.68% untuk *Support Vector Machine* (SVM) dan *Naïve Bayes* dengan akurasi 88.54%, dengan menggunakan dataset teks komentar Instagram sebanyak 776 *record* yang sudah dilakukan teks *preprocessing*.

Kata Kunci: Klasifikasi, Komentar Instagram, Keluhan, SVM, *Naïve Bayes*, Teknik SMOTE

**Abstract.** *Customer satisfaction is one of the things expected by a company when the product produced has been marketed, both in the form of goods and services. How to complain through customer service is very diverse, lately not only by telephone, customers submit their suggestions or complaints. Customers can submit their suggestions or complaints via e-mail or e-mail or forums in the virtual world that are made by product-producing companies to accommodate a variety of complaints, suggestions, and direct criticism from consumers, especially social media, who are free to express their opinions on shipping services. they use. Instagram is a social media that is more inclined to images and on the other hand has text captions and comments, from the above problems trying to make a research for customer complaints of users of goods delivery services on an Instagram account shipping service company. From the background of the problem, the researchers tried to solve the problem for text mining classifiers by using the Support Vector Machine (SVM) and Naïve Bayes methods and using the SMOTE technique with the usual processes for text mining so that they could produce 69.68% accuracy for Support Vector Machine (SVM) and Naïve Bayes with an accuracy of 88.54%, using the Instagram comment text dataset of 776 records that have been done with preprocessing text.*

*Key words: Classification, Comments Instagram, Complaints, SVM, Naïve Bayes, SMOTE Technique.*

## PENDAHULUAN

Penggunaan sosial media di Indonesia sangat beragam baik laki-laki dan perempuan mulai dari kalangan anak-anak, remaja, dewasa, dan orang tua, menurut Layanan Internet Indonesia *Provider Association* (APJII), jumlah Internet pengguna pada 2013 hingga 2015 meningkat dari 82 juta menjadi 139 juta orang. Sementara itu, sekitar 75,47% perusahaan menggunakan media sosial Instagram (Sabirin & Setiawati, 2017). Persaingan usaha yang sangat ketat saat ini mengharuskan perusahaan harus selalu mencari strategi penjualan baru agar bisa bersaing dengan perusahaan-perusahaan yang lainnya, salah satu aktivitas sebuah perusahaan adalah menjual untuk mendapatkan keuntungan, untuk mencapai tujuan tersebut sebuah perusahaan dituntut agar dapat memperkenalkan serta memasarkan produk yang dijualnya secara cepat tetapi tidak banyak *resource* yang terbuang (Hartini, 2016).

Kepuasan konsumen atau kepuasan pelanggan adalah salah satu hal yang diharapkan suatu perusahaan ketika produk yang dihasilkan telah dipasarkan, baik berupa barang maupun jasa. Ditambah lagi dengan persaingan ketat antar perusahaan yang menghasilkan produk serupa membuat perusahaan berlomba-lomba menghasilkan produk yang sesuai dengan keinginan konsumen. Akan tetapi, tak selamanya produk yang sampai di tangan pelanggan benar-benar bebas dari masalah, dengan perkembangan jual beli online melalui berbagai macam media baik pembelian melalui situs *e-commerce* atau melalui sosial media sangat memudahkan, menurut (Widagdo, 2016) perkembangan *e-commerce* tersebut sangat menarik untuk dibahas karena seiring dengan berjalannya waktu, industri telekomunikasi akan semakin berkembang baik dari segi jangkauan layanan maupun kecepatan koneksi internet. Dukungan sistem pembayaran yang semakin mudah, maka transaksi elektronik atau kegiatan *e-commerce* akan semakin berkembang.

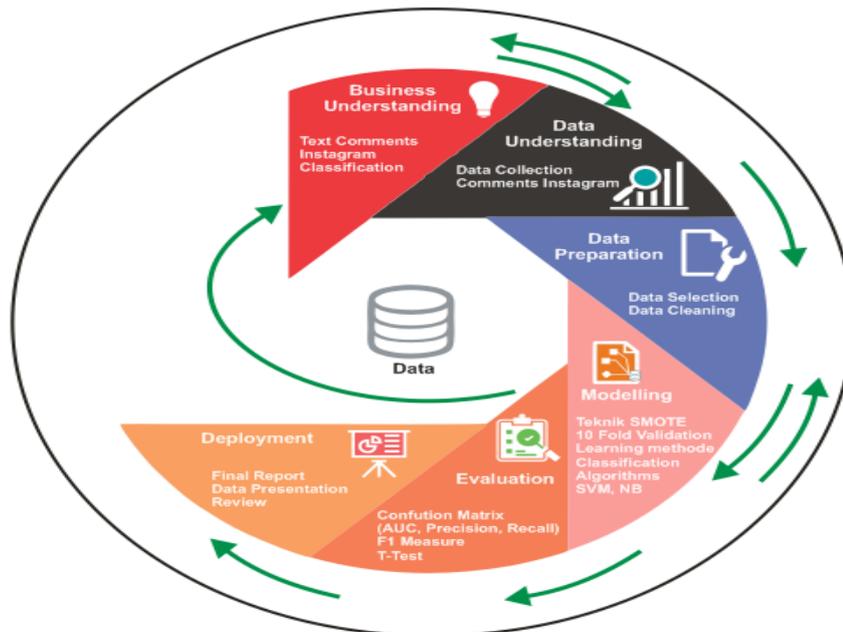
Semakin banyak keluhan yang diberikan oleh pelanggan membutuhkan adanya perhatian ekstra bagi setiap perusahaan yang menerima keluhan untuk memperbaiki dalam pembuatan produk atau jasa (Indriyani & Mardiana, 2016). Seringkali pelanggan menemukan masalah dalam penggunaan produk, baik dikarenakan kesalahan produksi atau kesalahan cara penggunaan selain dari cara pengiriman barang yang relatif lebih sering terjadi masalah yang dialami konsumen sehingga kepercayaan konsumen terhadap jasa pengiriman barang juga sangat berpengaruh karena produk yang dibeli oleh konsumen pasti sangat diharapkan cepat sampai ditangan konsumen.

Cara pengaduan melalui *customer service* sangat beragam, belakangan ini tak hanya melalui telepon saja pelanggan menyampaikan saran atau keluhannya. Pelanggan bisa menyampaikan saran atau keluhannya melalui surat elektronik atau e-mail maupun forum-forum di dunia maya yang memang dibuat perusahaan penghasil produk untuk menampung beragam keluhan, saran, dan kritik langsung dari konsumen terlebih sosial media yang bebas mengutarakan pendapat mereka terhadap jasa pengiriman yang mereka pakai. Setelah melalui langkah ini, perusahaan yang diwakili oleh *customer service* mengumpulkan dan menganalisa keluhan dari pelanggan atau konsumen untuk mengambil tindakan pada langkah selanjutnya, yang tentunya sesuai dengan permintaan konsumen atau pelanggan (Dewi, 2018).

Ada beberapa penelitian yang dilakukan tentang keluhan konsumen melalui *text mining* klasifikasi diantara penelitian yang dilakukan oleh Prita Dellia dan Aris Tjahyanto tentang klasifikasi keluhan pajak pada twitter tahun 2017 (Dellia & Tjahyanto, 2017). penelitian lainnya mengenai teks klasifikasi twitter untuk keluhan pemerintahan kota bandung (Laksana & Purwarianti, 2015). Metode Klasifikasi banyak sekali membantu mengidentifikasi keputusan berdasarkan model pembelajaran terawasi (Irfansyah, 2016). Penulisan penelitian ini sendiri menggunakan *text mining* klasifikasi untuk keluhan konsumen pada sosial media instagram dengan mengambil teks komentar pada salah satu akun instagram jasa pengiriman barang dengan menggunakan metode SVM (*Support Vector Machine*) dan Naive Bayes dan juga ditambahkan teknik SMOTE sehingga dapat membedakan antara akurasi yang memakai teknik SMOTE atau tidak yang dihasilkan pada penelitian ini.

## METODE

Metodologi penelitian yang digunakan dalam penelitian ini dengan menggunakan metode *Cross-Standard Industry for Data Mining* (CRISP-DM), Metode penelitian CRISP-DM terdiri dari 6 langkah (Zhu, 2017), yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, *Deployment*.



Gambar 1. Proses Models CRISP-DM

Metode algoritma yang digunakan pada penelitian ini ada algoritma klasifikasi diantaranya:

### **Algoritma SVM (Support Vector Machine)**

Algoritma SVM (*Support Vector Machine*) adalah metode untuk klasifikasi data linier dan nonlinier. Singkatnya, SVM adalah algoritma yang berfungsi menggunakan pemetaan nonlinear untuk mengubah data pelatihan asli menjadi dimensi yang lebih tinggi. (Han, Kamber, & Jian, 2012).

### **Algoritma Naive Bayes**

Algoritma Naive Bayes memberi kita cara untuk menggabungkan probabilitas sebelumnya dan probabilitas kondisional dalam formula tunggal, yang dapat kita gunakan untuk menghitung probabilitas masing-masing klasifikasi yang mungkin pada gilirannya. Setelah selesai kita memilih klasifikasi dengan nilai terbesar (Bramer, 2007).

### **Teknik SMOTE**

Teknik *SMOTE* (*Synthetic Minority Over-sampling Technique*) merupakan metode yang populer diterapkan dalam rangka menangani ketidak seimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan *dataset* dengan cara membuat instance baru dari minority class dengan pembentukan *convex* kombinasi dari *instances* yang saling berdekatan. Secara efektif menggambarkan garis antara *minority point* di ruang fitur dan sampel di sepanjang garis ini (Arwan et al., 2018)

Ada beberapa yang dapat dilihat hasil dari penelitian *text mining* klasifikasi ini, antara lain sebagai berikut:

### **Confusion Matrix (Accuracy)**

*Confusion Matrix* merupakan tabel pencatat hasil kerja klasifikasi. Tabel *confusion matrix* berisikan empat kemungkinan keluaran sebagai bahan acuan dalam membandingkan antara kejadian yang sebenarnya (aktual) dengan kejadian yang terprediksi. Metode ini menggunakan tabel matriks seperti pada Tabel 1 (Essra, Rahmadani, & Safriadi, 2016).

Tabel 1. Model Confusion matrix

		Aktual	
		<i>Complaint</i>	Not <i>Complaint</i>
Prediksi	<i>Complaint</i>	TP	TN
	Not <i>Complaint</i>	FP	FN

dimana:

*True Positive* (TP) sama dengan jumlah data *complaint* yang diprediksi *complaint*.

*True Negative* (TN) sama dengan jumlah data *complaint* yang diprediksi *not complaint*.

*False Positive* (FP) sama dengan jumlah data *not complaint* yang diprediksi *complaint*.

*False Negative* (FN) sama dengan jumlah data *not complaint* yang diprediksi *not complaint*.

Jumlah data yang diklasifikasikan benar, dapat digunakan untuk menghitung performansi akurasi dan *Detection Rate*, sedangkan jumlah data yang diklasifikasikan salah, dapat digunakan untuk menghitung performansi *False Positive Rate* dan *True Negative Rate*. Akurasi dapat *dihitung* menggunakan rumus:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Dimana:

TP : Jumlah kasus positif yang diklasifikasikan sebagai positif

FP : Jumlah kasus negatif yang diklasifikasikan sebagai positif

TN : Jumlah kasus negatif yang diklasifikasikan sebagai negatif

FN : Jumlah kasus positif yang diklasifikasikan sebagai negatif

### ROC Curve

Kurva ROC adalah tool dua dimensi yang dipergunakan untuk menilai hasil kinerja klasifikasi yang menggunakan dua class sebagai keputusannya, objek dipetakan ke salah satu elemen dari himpunan pasangan, positif atau negatif. Pada kurva ROC, TP rate diplot pada sumbu Y dan FP rate diplot pada sumbu X (Frastian, Hendrian, & V.H. Valentino, 2018).

### Precision & Recall

Evaluasi menggunakan nilai precision dan recall dilakukan untuk mengetahui tingkat relevansi dan ketepatan sistem dalam melakukan pencarian informasi yang diminta oleh pengguna (Novianti, Setiawan, & Kusumawardani, 2015).

$$Precision = x \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

### F-Measure

Merupakan nilai yang didapatkan dari pengukuran *precision* dan *recall* antara class hasil *cluster* dengan class sebenarnya yang terdapat pada data masukan.

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

## HASIL DAN PEMBAHASAN

### *Business Understanding*

Pemahaman bisnis terhadap apa yang akan di bahas dalam penelitian ini sebagai memahami proses yang akan di dihasilkan baik bagi organisasi maupun individu. Apa yang akan dibahas pada penelitian ini sebagai memahami organisasi dalam untuk meningkat kualitas organisasi, pada tahapan ini peneliti membuat suatu pertanyaan apa tujuan dari penelitian *text mining* ini? yaitu untuk meningkatkan pelayanan pada jasa pengiriman barang dalam menanggapi keluhan pelanggannya.

Berdasarkan ini, pendekatan model klasifikasi pada komentar instagram yang akan digunakan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) serta menambahkan teknik *SMOTE* untuk meningkatkan akurasi dari metode klasifikasi yang digunakan dalam *dataset* teks komentar instagram.

#### **Data Understanding**

Tahap kedua *CRISP-DM* adalah data *understanding* dibutuhkan adanya pemahaman mengenai data yang akan digunakan dalam proses penelitian. Data yang digunakan untuk eksperimen penelitian ini dikumpulkan melalui data komentar instagram yang berasal dari sosial media instagram mengenai keluhan pelanggan pada salah satu akun jasa pengiriman barang dengan menggunakan *Instagram API Python* memakai Bahasa python yang ada di Github dengan hanya mengambil data *local text* komentar instagram, kemudian data local di labelkan hasil penyeleksian dataset yang dibuat pada *Microsoft Excel* untuk diolah dalam pengujian data.

#### **Data Preparation**

*Data Preparation* yaitu melakukan tahap persiapan data untuk proses *text mining* pengambilan data tersebut dari komentar instagram akun jasa pengiriman yaitu @jne\_id dari periode September 2018 sampai November 2018. Struktur data yang diambil dari instagram terdiri dari beberapa kolom yaitu nama akun dan teks komentar yang digunakan pada batasan masalah ini hanya teks komentar dan dilakukan penambahan *field* baru yaitu status untuk dijadikan sebagai *class* atau label. Isi dari instagram memiliki banyak informasi yang berkaitan keluhan pelanggan pengguna sosial media Instagram yang ditunjukkan kepada akun instagram jasa pengiriman yaitu @jd\_id. Sebelum di hitung menggunakan model yang diusulkan data tersebut di *cleaning* agar data tersebut sesuai dengan perhitungan metode dengan menghilangkan data noise dan data yang tidak konsisten, tahapan *data preparation* sebagai berikut:

#### **@Anotation Removal**

Teks diurai berdasarkan *white space*. Dalam proses ini, semua anotasi yang terkandung dalam komentar instagram akan dihilangkan dan mengubah seluruh huruf kapital menjadi huruf kecil .

Tabel 2. Proses *Preprocessing @Anotation Removal*

Teks sebelum proses @Anotation Removal	Teks setelah proses @Anotation Removal
@adhyepatie87 tanpa customer @jne_id tidak akan ada di muka bumi ini. Tolong ya dengarkan keluh kesah kami. Setiap pekerjaan pasti ada resikonya. Dan saat inilah JNE harus menghadapi itu	tanpa customer tidak akan ada di muka bumi ini. tolong ya dengarkan keluh kesah kami. setiap pekerjaan pasti ada resikonya. dan saat inilah jne harus menghadapi itu.

Keterangan : Blok kata yang akan diproses.

#### **Transformation Remove URL**

Hasil dari *@anotation removal* akan di lanjutkan ke proses *Transformation Remove URL*, dalam proses ini link atau URL yang terkandung pada komentar instagram akan dihilangkan.

Tabel 3. Proses *Preprocessing Transformation Remove URL*

Teks sebelum proses <i>Tokenize Remove HTTP/HTTPS</i>	Teks setelah proses <i>Tokenize Remove HTTP/HTTPS</i>
Anda tidak perlu blokir saya , saya bisa buat 1000 akun linkedin kalau perlu. yang cs anda perlu lakukan adalah tutup akun linkedin jne atau tingkatkan kinerja kurir anda. Resi 081800034760918 atas nama devi fitriani kenapa belum sampai?tolong diganti semua saja kurir anda yang area cirebon. kok bisa sampai pelanggan saya kapok sama kurir jne disana. Saya mau ada info secepatnya <a href="https://www.jne.co.id/id/tracking/trace">https://www.jne.co.id/id/tracking/trace</a>	anda tidak perlu blokir saya , saya bisa buat 1000 akun linkedin kalau perlu. yang cs anda perlu lakukan adalah tutup akun linkedin jne atau tingkatkan kinerja kurir anda. resi 081800034760918 atas nama devi fitriani kenapa belum sampai?tolong diganti semua saja kurir anda yang area cirebon. kok bisa sampai pelanggan saya kapok sama kurir jne disana. saya mau ada info secepatnya

Keterangan : Blok kata yang akan diproses.

#### *Tokenization (Regex)*

Hasil dari *Transformation Remove URL* dilanjutkan oleh proses *Tokenization (Regex)* yaitu semua kata yang ada didalam tiap dokumen dikumpulkan dan dihilangkan tanda *backslash*, serta dihilangkan jika terdapat simbol, karakter khusus atau apapun yang bukan huruf.

Tabel 4. Proses *Preprocessing Tokenization (Regex)*.

Teks sebelum proses <i>Tokenization</i>	Teks setelah proses <i>Tokenization</i>
Acara ultahnya sudah selesai kan tolong sekarang focus urus pengiriman yang sudah tidak jelas ! barang saya dari tanggal 5 on transit terus di makassar dan saya butuh barang saya untuk hari rabu!!!tolong @jne_id di perhatikan lagi!!!dan tolong direspon!!!!!!!!	acara ultahnya sudah selesai kan tolong sekarang focus urus pengiriman yang sudah tidak jelas barang saya dari tanggal on transit terus di makassar dan saya butuh barang saya untuk hari rabu tolong di perhatikan lagi dan tolong direspon

Keterangan : Blok kata yang akan diproses.

#### *Transformation Not Negative*

Dari hasil *Normalization Indonesian Slang* selanjutnya dilakukan proses *transformation not negative*. Dalam proses ini kata-kata yang memiliki nilai negatif akan di dirubah menjadi kalimat tidak negatif.

Tabel 5. Proses *Preprocessing Transformation Not Negative*

Teks sebelum proses <i>transformation not negative</i>	Teks setelah proses <i>transformation not negative</i>
astaga jne kapan kalian mau kirim paket kami dari batam keterlaluhan kalian kalau memang tidak sanggup kirim info ya dari awal ke pengirim jangan semua paket diterima cuma buat ditimbun di sorting center	astaga jne kapan kalian mau kirim paket kami dari batam keterlaluhan kalian kalau memang tidak_sanggup kirim info ya dari awal ke pengirim jangan semua paket diterima cuma buat ditimbun di sorting center

Keterangan : Blok kata yang akan diproses.

### Indonesian Stemming

Hasil dari transformation *not negative* akan dilakukan proses *stemming* yaitu mengubah kata berimbuhan menjadi kata dasar menggunakan *indonesian stemming* untuk komentar instagram berbahasa.

Tabel 6. Proses *Preprocessing Indonesia Stemming*

Teks sebelum proses <i>Indonesia stemming</i>	Teks setelah proses <i>Indonesia stemming</i>
astaga jne kapan kalian mau kirim paket kami dari batam keterlaluhan kalian kalau memang tidak sanggup kirim info ya dari awal ke pengirim jangan semua paket diterima cuma buat ditimbun di sorting center	astaga jne kapan kalian mau kirim paket kami dari batam keterlaluhan kalian kalau memang tidak_sanggup kirim info ya dari awal ke pengirim jangan semua paket terima cuma buat timbun di sorting center

Keterangan : Blok kata yang akan diproses.

### Indonesian Stopword Removal

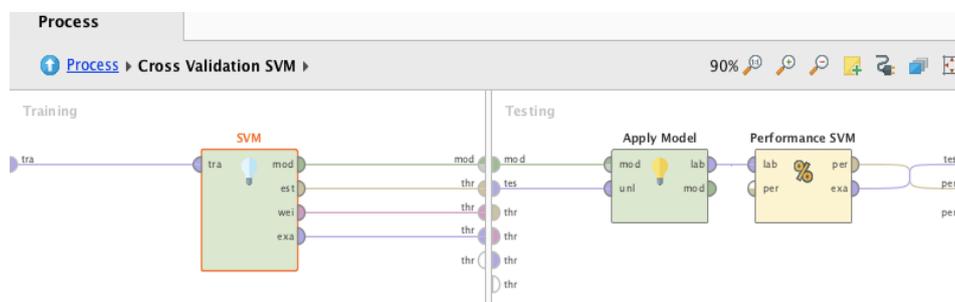
Hasil dari *indonesian stemming* akan di lanjutkan ke proses *Indonesian stopword removal*, dalam proses ini kata-kata yang tidak relevan akan dihapus, seperti kata tetapi, untuk, dengan, yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan *sentiment*.

Tabel 7. Proses *Preprocessing Indonesian Stopword Removal*

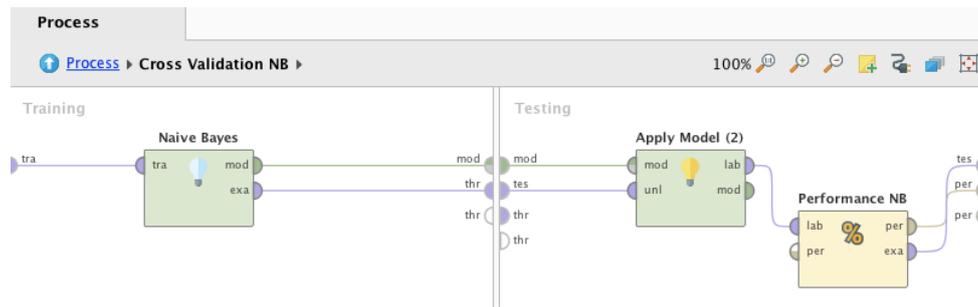
Teks sebelum proses <i>Indonesia stemming</i>	Teks setelah proses <i>Indonesian stopword removal</i>
astaga jne kapan kalian mau kirim paket kami dari batam keterlaluhan kalian kalau memang tidak_sanggup kirim info ya dari awal ke pengirim jangan semua paket terima cuma buat timbun di sorting center	astaga jne kirim paket batam tidak_sanggup kirim info ya pengirim paket terima timbun sorting center

### Modelling

Metode yang digunakan dalam penelitian ini yaitu metode klasifikasi *text mining* ini yaitu *Support Vector Machines (SVM)* dan *Naïve Bayes (NB)* sebagai pembandingnya, dilakukan perbandingan kedua metode pemilihan tersebut untuk diketahui metode pemilihan terbaik untuk diterapkan dengan menggunakan pengklasifikasi *Support Vector Machines (SVM)* dan *Naïve Bayes (NB)* karena merupakan teknik *machine learning* yang populer untuk klasifikasi teks serta memiliki performa yang baik pada banyak domain. Lihat gambar berikut:



Gambar 2. Desain Model Naïve Bayes



Gambar 3. Desain Model Support Vector Machine (SVM)

**Evaluation**

Pada tahap evaluasi ini hasil dari pengujian model yang dilakukan adalah mengklasifikasikan komentar instagram pada akun sebuah perusahaan jasa pengiriman yaitu @jne\_id menjadi *complaint* dan *notcomplaint* sebagai data teks komentar instagram menggunakan algoritma *Support Vector Machine* dan *Support Vector Machine* berbasis *Smote UpSampling*, Naïve Bayes dan Naïve Bayes berbasis *Smote UpSampling*, untuk mengetahui model terbaik pada masing-masing model berikut:

**Confusion Matrix SVM**

Hasil pengujian data komentar instagram bahasa Indonesia dengan data 776 record menggunakan metode *Support Vector Machine* (SVM) dan Naïve Bayes dapat dilihat pada table.

Tabel 8. Confusion Matrix Metode SVM

Accuracy: 69,68% +/- 3.71%			
	True Complaint	True NotComplaint	Class Precision
Pred. Complaint	222	4	98,23%
Pred. NotComplaint	332	550	62,36%
Class Recall	40,07%	99,28%	

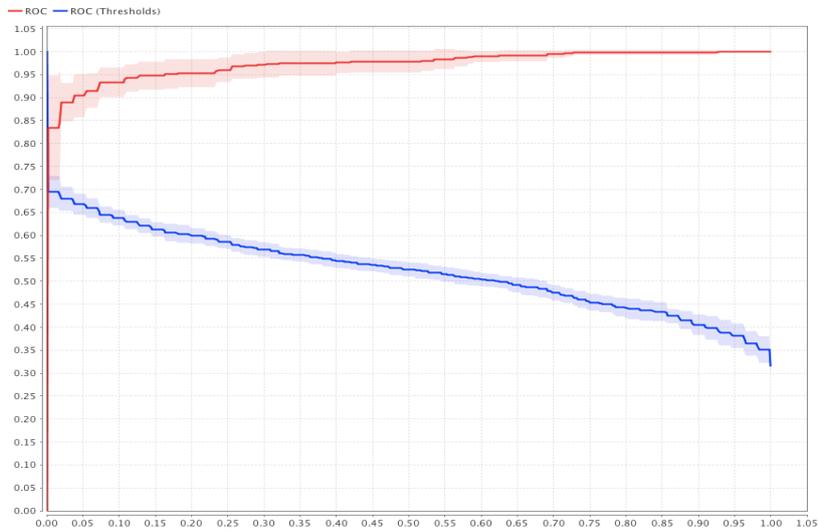
Tabel 9. Confusion Matrix Metode Naïve Bayes

Accuracy: 88,54% +/-2,38%			
	True Complaint	True NotComplaint	Class Precision
Pred. Complaint	461	34	93,13%
Pred. NotComplaint	93	520	84,83%
Class Recall	83,21%	93,86%	

**Kurva Receiver Operating Characteristic (ROC) Model SVM**

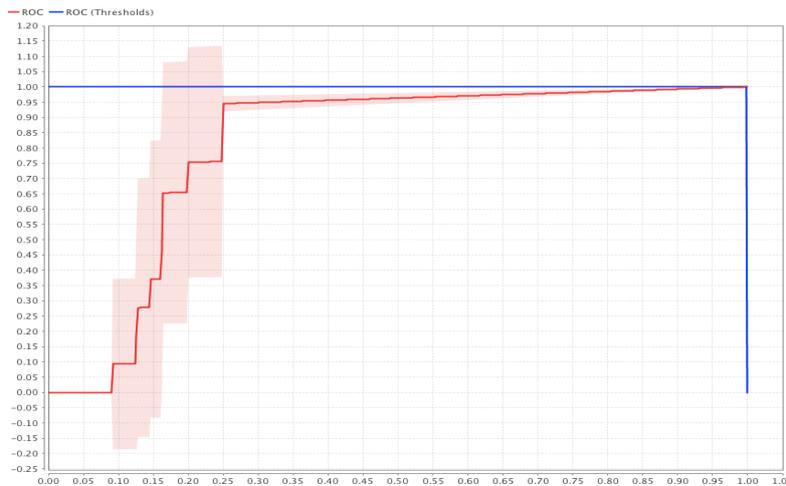
Hasil pengujian data teks komentar instagram bahasa Indonesia dengan data 776 menggunakan metode SVM dan Naïve Bayes menghasilkan:

Area Under Curve (AUC):0.973 +/- 0.015 (Micro average:0.973)



Gambar 4. Kurva ROC Support Vector Machine

Area Under Curve (AUC):0.500 +/- 0.000 (Micro average:0.500)



Gambar 5. Kurva ROC Naïve Bayes

**Precision Recall SVM dan Naïve Bayes**

Tabel 10. Precision dan Recall SVM

	SVM SMOTE Precision	SVM SMOTE Recall
pred. COMPLAINT	98.23%	40.07%
pred. NOT COMPLAINT	62.36%	99.28%

Tabel 11. Precision dan Recall Naïve Bayes

	NB SMOTE Precision	NB SMOTE Recall
pred. COMPLAINT	93.13%	83.21%
pred. NOT COMPLAINT	84.83%	93.86%

### F-Measure SVM dan Naïve Bayes

Tabel 12. Evaluasi F-Measure

	SVM	NB	SVM SMOTE	NB SMOTE
F-Measure	6.09%	51.54%	76.66%	89.14%

### Evaluasi Pengujian T-Test

Tabel 13. T-Test SVM dan Naïve Bayes

A	B	C
-	0.697+/- 0.037	0.885 +/- 0.024
0.697 +/-0.037	-	0.000
0.885+/- 0.024	-	-

### Deployment

Berdasarkan hasil evaluasi dari proses pengujian model algoritma *Support Vector Machine* dan *Naïve Bayes* tanpa fitur serta *Support Vector Machine* dan *Naïve Bayes* dengan menggunakan teknik SMOTE, hasil pengujian model tertinggi dari seluruh hasil pengujian algoritma adalah *Naïve Bayes* berbasis SMOTE. Maka dari itu bobot yang akan digunakan dalam pemodelan aplikasi berdasarkan dari hasil pengujian dari algoritma *Naïve Bayes* berbasis teknik SMOTE.

## PENUTUP

### Simpulan

Pada penelitian ini telah berhasil dibangun model untuk melakukan klasifikasi keluhan pelanggan jasa pengiriman pada akun instagram. Secara garis besar, terdapat dua jenis model yang dibangun dengan dua pendekatan yang berbeda yaitu *Support Vector Machine* dan *Naive Bayes*. Berdasarkan hasil eksperimen, model *Support Vector Machine* memiliki akurasi yang lebih rendah dengan perbedaan yang cukup signifikan jika dibandingkan dengan model yang dihasilkan dari algoritma *Naive Bayes* berbasis Teknik SMOTE dengan menghasilkan akurasi 88,54%. Hal tersebut kemungkinan disebabkan algoritma *Support Vector Machine* yang digunakan dalam percobaan ini adalah algoritma *Support Vector Machine* dengan linear kernel.

Linear kernel pada *Support Vector Machine* (SVM) hanya berjalan optimal untuk kasus klasifikasi biner (*binary classification*), sedangkan data keluhan yang digunakan dalam percobaan ini memiliki lebih dari satu jenis kelas label sehingga dikelaskan sebagai *multiclass classification*. *Support Vector Machine* (SVM) akan berjalan optimal pada *multiclass classification* dengan menggunakan kernel yang didesain untuk data multidimensi. Hal tersebut diluar cakupan dari percobaan ini dan akan menjadi bagian dalam penelitian selanjutnya.

## Saran

Adapun saran yang ingin peneliti sampaikan sesuai dengan kebutuhan pengolahan data adalah, sebagai berikut :

1. Menggunakan algoritma pengklasifikasian yang lain seperti k-nn sehingga bisa dilakukan penelitian yang berbeda dari yang sudah ada.
2. Menggunakan metode pemilihan fitur yang lain seperti *Chi Square*, *Mutual Information Part of Speech (POS)*, dan lain-lain agar hasilnya bisa dibandingkan.
3. Memperbanyak *dataset* yang digunakan sehingga penelitian dapat lebih spesifik dan akurat.

## DAFTAR PUSTAKA

- Arwan, Ardina, V., Ariana, L. R., Samuel, F., Ramdani, D., Aditya, & Sukmana, E. A. (2018). Synthetic Minority Over-Sampling Technique (Smote) Algorithm For Handling Imbalanced Data.
- Bramer, M. (2007). *Principles of Data Mining*. Springer London.
- Dellia, P., & Tjahyanto, A. (2017). *Tax Complaints Classification on Twitter Using Text Mining*. 2(1).
- Dewi, R. N. (2018). *Model Text Mining Untuk Identifikasi Keluhan Pelanggan Produk Perusahaan Perangkat Lunak*.
- Essra, A., Rahmadani, & Safriadi. (2016). Analisis Information Gain Attribute Evaluation Untuk Klasifikasi Serangan Intrusi. *Journal of Information System Development*, 2(2), 9–14.
- Frastian, N., Hendrian, S., & V.H.Valentino. (2018). *Implementasi komparasi algoritma klasifikasi menentukan kelulusan mata kuliah algoritma universitas budi luhur*. (August), 0–8. <https://doi.org/10.13140/RG.2.2.31821.74721>
- Han, J., Kamber, M., & Jian, P. (2012). Data Mining: Concepts and Techniques. In *San Francisco, CA, itd: Morgan Kaufmann*. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Hartini, S. (2016). *Efektifitas Endorsment Pada Media Sosial Instagram Pada Produk Skin Care*. 3(1), 43–50.
- Indriyani, S., & Mardiana, S. (2016). *Pengaruh Penanganan Keluhan ( Complaint Handling ) Terhadap Kepercayaan Dan Komitmen Mahasiswa Pada Perguruan Tinggi Swasta Di Bandar Lampung*. 2(01), 1–13.
- Irfansyah, P. (2016). Kajian Komparasi Penerapan Algoritma Data Mining (C4.5, Bayesian Classifier Dan Neural Network) Dalam Menentukan Promosi Jabatan. *Prosiding Seminar Nasional*, 53–67.
- Laksana, J., & Purwarianti, A. (2015). Indonesian Twitter text authority classification for government in Bandung. *Proceedings - 2014 International Conference on Advanced Informatics: Concept, Theory and Application, ICAICTA 2014*, 129–134. <https://doi.org/10.1109/ICAICTA.2014.7005928>
- Novianti, K. D. P., Setiawan, N. A., & Kusumawardani, S. S. (2015). *Peningkatan Nilai Recall dan Precision pada Penelusuran Informasi Pustaka Berbasis Semantik ( Studi Kasus : Sistem Informasi Ruang Referensi Jurusan Teknik Elektro dan Teknologi Informasi UGM)*. 9–10.
- Sabirin, & Setiawati, C. I. (2017). The driving factors of instagram utilization for marketing efforts in promoting student owned online store. *Proceedings - 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016*, 64–69. <https://doi.org/10.1109/ISEMANTIC.2016.7873811>
- Widagdo, P. B. (2016). *Perkembangan Electronic Commerce ( E- Commerce ) di Indonesia*. (December), 10.
- Zhu, X. (2017). *Agile mining : a novel data mining process for industry practice based on Agile Methods and visualization*.