

Analisis Komentar Potensial pada *Social Commerce* Instagram Menggunakan TF-IDF

Rizqa Luviana Musyarofah¹, Ema Utami², Suwanto Raharjo³

^{1,2}Magister Teknik Informatika, ³Informatika

^{1,2}Universitas Amikom Yogyakarta, ³IST Akprind Yogyakarta
Yogyakarta, Indonesia

e-mail: ¹rizqa.1074@students.amikom.ac.id, ²ema.u@amikom.ac.id, ³wa2n@nrar.net

Diajukan: 11 Februari 2020; Direvisi: 28 Februari 2020; Diterima: 2 Maret 2020

Abstrak

Komentar di Instagram sangat berharga, informatif dan sangat membantu. Bagi penjual komentar adalah fitur yang menunjukkan respons pengguna Instagram terhadap produk yang ditawarkan, dan melalui fitur komentar penjual dapat menemukan pelanggan yang potensial. Manfaat tersebut diperoleh apabila penjual melakukan analisis pada komentar di toko Instagram-nya. Sangat dimungkinkan untuk menganalisis secara manual apabila data komentar pada tokonya berjumlah sedikit namun apabila komentar yang dimiliki banyak maka akan lebih cepat apabila menggunakan sistem. Banyaknya spam dapat mengganggu informasi yang ada pada komentar, sehingga tidak menjamin banyaknya komentar pada sebuah posting-an maka banyak pula yang ingin membeli produk tersebut. Oleh karena itu dibutuhkan sistem yang bisa memfilter komentar agar penjual dapat menemukan pelanggan yang tepat untuk produknya. Penelitian ini menggunakan algoritma TF-IDF untuk mengklasifikasikan komentar ke dalam 2 kelas (potensial dan tidak potensial) dan memperoleh akurasi sebesar 80%, presisi 0,76 dan recall 0,94. Berdasarkan hasil penelitian pada 294 komentar, 27% di antaranya adalah komentar tidak potensial. Kata yang menunjukkan minat beli seseorang adalah "berapa", "kak", "ada", dan "tidak", sedangkan kata dominan pada komentar tidak potensial adalah kata "mention" yang menunjukkan aktivitas mention.

Kata kunci: Analisis sentimen, Social commerce, TF-IDF, Text mining, Komentar Instagram.

Abstract

Comments on Instagram are very valuable, informative and very helpful. For the seller the comment is a feature that shows the response of Instagram users to the product offered, and through the comments feature the seller can find potential customers. These benefits are obtained when the seller analyzes the comments on his Instagram store. It is possible to analyze manually if the comment data in the shop is small but if there are many comments, it will be faster to use the system. The amount of spam can disrupt the information in the comments, so it does not guarantee the number of comments on a post so many people want to buy the product. Therefore, a system that can filter comments is needed so that the seller can find the right customer for his product. This study uses the TF-IDF algorithm to classify comments into 2 classes (potential and non-potential) and obtain an accuracy of 80%, 0.76 precision and 0.94 recall. Based on the results of the study on 294 comments, 27% of them are non-potential comments. The words that indicate someone's buying interest are "berapa", "kak", "ada" and "tidak", while the dominant word in comments with no potential is the word "mention" which indicates the activity of mention.

Keywords: Sentiment Analysis, Social commerce, TF-IDF, Text mining, Instagram comment.

1. Pendahuluan

Social Commerce (*s-commerce*) merupakan pengembangan dari *e-commerce*, kegiatan perdagangan elektronik yang memanfaatkan media sosial dan teknologi *web 2.0* [1]. Pada platform *s-commerce* konsumen dapat berinteraksi satu sama lain dan mencari informasi yang diperlukan untuk membantu mereka dalam membuat keputusan pembelian. Berdasarkan survei We Are Social dan Hootsuite pada tahun 2018, Instagram adalah platform media sosial populer urutan ke 4 di Indonesia dengan total pengguna aktif sebanyak 53 juta [2]. Kondisi ini di dimanfaatkan oleh para pebisnis Usaha Mikro Kecil Menengah (UMKM) untuk menggunakan Instagram sebagai platform *s-commerce*. Hasil survei Instagram

kepada 502 pemilik akun bisnis di Instagram, 66% responden mengatakan bahwa Instagram membantu dalam menemukan pelanggan baru [3].

Menurut [4], komentar di Instagram sangat berharga, informatif, dan sangat membantu. Bagi penjual komentar adalah fitur yang menunjukkan respons pengguna Instagram terhadap produk yang ditawarkan. Selain itu melalui komentar pelanggan dan penjual bisa saling berinteraksi hingga melakukan *closing* penjualan. Manfaat komentar di toko Instagram bagi pembeli adalah untuk mencari informasi yang diungkapkan oleh pengguna lain pada fitur komentar menjadi pertimbangan dalam membeli produk tersebut. Hal yang dapat mengganggu informasi pada komentar di Instagram adalah *spam*. *Spam* adalah sebuah pesan yang tidak diinginkan, tidak berhubungan dengan konten yang diunggah atau pesan yang sama dan dikirim oleh seorang pengguna lebih dari satu kali. Hal ini dibuktikan oleh penelitian [5] dari 17.007 komentar yang ada di akun Instagram artis Indonesia 10,719 di antaranya adalah *spam*. Pada penelitian selanjutnya hasil analisis komentar diklasifikasikan ke dalam dua kelas yaitu komentar potensial dan komentar tidak potensial. *Spam* termasuk ke dalam kelas komentar tidak potensial.

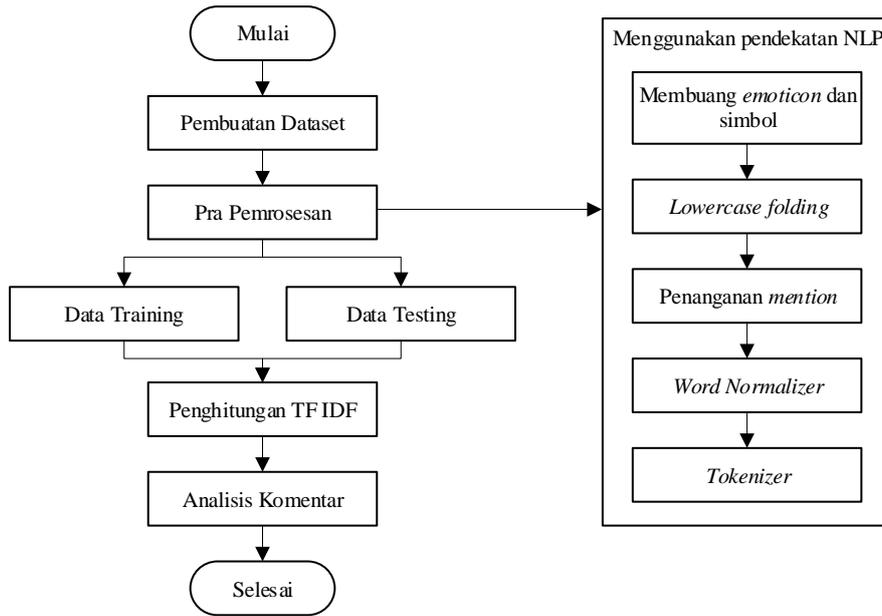
Pada penelitian ini peneliti melakukan analisis komentar potensial pada akun toko Instagram. Komentar potensial yang dimaksud adalah komentar pengguna yang tertarik dengan produk yang ditawarkan oleh akun toko Instagram tersebut dan bertujuan untuk melakukan pembelian. Indikator minat beli dapat dijelaskan oleh komponen-komponen sebagai berikut: 1) Tertarik untuk mencari informasi tentang produk 2) Mempertimbangkan untuk membeli 3) Tertarik untuk mencoba 4) Untuk mengetahui produk 5) Ingin memiliki [6]. Penelitian mengenai minat beli pada Instagram mengatakan bahwa sikap positif dan negatif dapat membentuk minat seseorang dalam belanja produk. Sikap positif ditunjukkan dari adanya ketertarikan untuk mencari informasi terhadap produk yang dijual dan setelah itu melakukan pembelian produk melalui Instagram. Sedangkan, sikap negatif dapat ditunjukkan dari tidak adanya ketertarikan untuk mencari informasi tentang produk yang diiklankan dan tidak berminat untuk membeli produk tersebut [7].

Beberapa penelitian sebelumnya mengatakan bahwa ada hubungan yang positif antara komentar dengan niat pembelian [8], [9]. Menurut [8] keakraban, kehadiran sosial, kepercayaan, dan pencarian informasi melalui *s-commerce* memiliki pengaruh positif langsung pada niat pembelian. Penelitian tentang sentimen analisis pada media sosial sudah pernah dilakukan oleh peneliti sebelumnya. Penelitian [10] menganalisis komentar di Instagram untuk mengetahui sentimen publik pada sebuah Universitas, dengan metode klasifikasi menggunakan Naive Bayes dan mendapatkan akurasi sebesar 76,5%. Pada penelitian selanjutnya, algoritma pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk menganalisis komentar potensial. Metode TF-IDF adalah cara untuk memberi bobot pada hubungan suatu kata (term) dengan suatu dokumen. Metode ini menggabungkan dua konsep untuk kata dalam dokumen tertentu dan frekuensi terbalik dari dokumen yang mengandung kata tersebut. Algoritma ini berhasil digunakan untuk pengambilan artikel yang relevan pada situs berita *online* menggunakan kata kunci pencarian, dengan nilai *recall* 1 dan nilai rata-rata presisi adalah 0,5[11].

Tujuan dari penelitian ini adalah menganalisis komentar pada akun toko Instagram dan mengklasifikasikannya ke dalam 2 kelas (potensial dan tidak potensial). Melalui komentar potensial didapatkan akun pelanggan yang potensial, sehingga penjual dapat mengikuti akun pengguna dan menawarkan produk secara langsung. Hal tersebut membuat aktivitas *marketing* melalui Instagram lebih tepat sasaran. Melalui analisis komentar penjual dapat memperoleh respons pengguna pada produk yang telah di-*posting*. Permasalahan dalam menganalisis komentar pada Instagram adalah banyaknya penggunaan singkatan dan bahasa yang tidak baku sehingga sulit dimengerti. Oleh karena itu, pendekatan *Natural Language Processing* (NLP) diperlukan untuk memperbaiki bahasa pada komentar sehingga komentar lebih mudah dipahami oleh sistem.

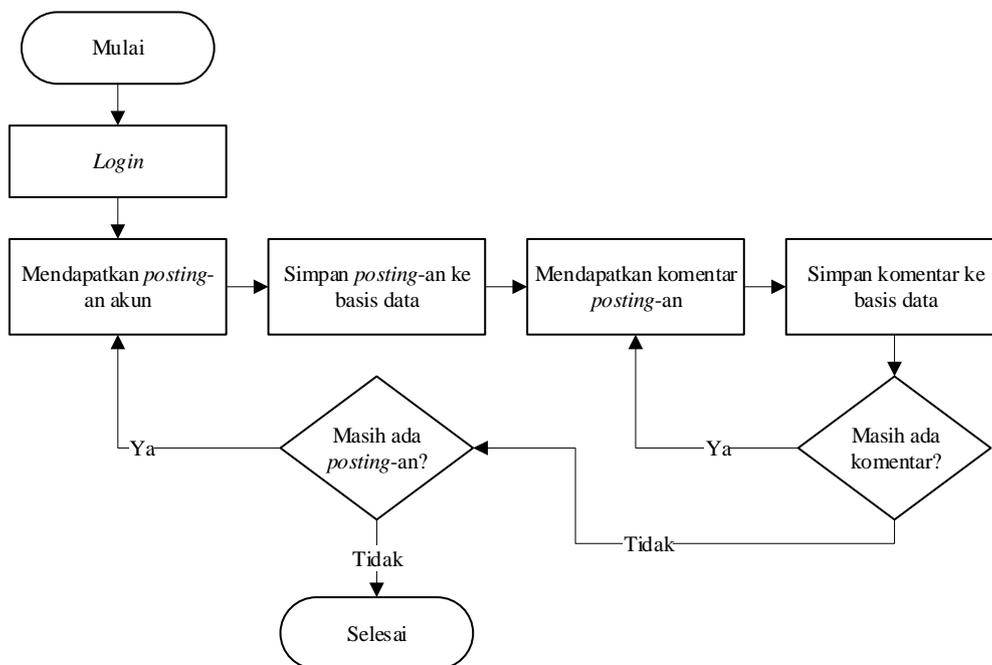
2. Metode Penelitian

Alur penelitian dalam analisis komentar potensial pada *social commerce* Instagram menggunakan TF-IDF dijelaskan pada Gambar 1.



Gambar 1. Alur penelitian.

Tahap awal penelitian ini adalah pembuatan *dataset*, yaitu menggunakan *dataset* komentar dari 10 akun toko di Instagram dengan 10 kategori produk jual yang berbeda. Akun toko ditentukan berdasarkan hasil pencarian produk menggunakan *hashtag* nama kategori produk. *hashtag* adalah tagar yang dapat ditambahkan pada keterangan atau komentar di pos Instagram, pos yang menggunakan tagar akan terlihat pada halaman tagar dan jika diklik oleh pengguna akan mengarah ke *posting*, gambar atau komentar serupa yang terkait dengan produk, layanan atau bahkan *merk* [9]. Akun toko Instagram yang dipilih adalah akun yang tidak dikunci dan memiliki jumlah *followers* minimal 50.000 dan jumlah *posting-an* minimal 20. Pengambilan *dataset* komentar menggunakan *tool scrapping* yang dibuat mandiri menggunakan bahasa *python*. Alur kerja sistem *scrapping* dijelaskan pada Gambar 2.



Gambar 2. Proses pengambilan komentar menggunakan *tool scrapping*.

Pada setiap akun akan dipilih *posting*-an yang mengandung minimal 5 komentar, jika kurang dari batas minimal maka *posting*-an akan diabaikan. Jenis *posting*-an yang digunakan adalah foto bukan video. Komentar yang sama dan dibuat dari pengguna yang sama pada satu *posting*-an hanya akan diambil salah satu. Komentar disimpan dalam basis data MySQL agar lebih mudah dikelola. Hasil *scrapping* komentar dilabeli secara manual oleh 8 responden. Responden terdiri dari 5 responden sebagai pembeli atau pengguna Instagram yang memiliki pengalaman dalam membeli produk di Instagram dan 3 responden adalah penjual di Instagram. Sebanyak 300 komentar diberi label dalam 2 kelas, komentar potensial, dan komentar tidak potensial.

Pra pemrosesan menggunakan pendekatan *Natural Language Processing (NLP)* digunakan untuk memperbaiki bahasa yang ada pada komentar. NLP merupakan salah satu cabang ilmu AI yang berfokus pada pengolahan bahasa natural. Bahasa natural adalah bahasa yang secara umum digunakan oleh manusia dalam berkomunikasi satu sama lain [12]. Fitur NLP yang digunakan pada penelitian ini adalah:

- a. Membuang *emoticon*, angka dan simbol
 Komentar pada Instagram umumnya mengandung *emoticon*, pada penelitian ini *emoticon* dibuang karena makalah ini hanya membahas teks yang terkandung pada komentar. Selain itu, membuang angka dan membuang simbol seperti “~”, “\”, “!”, “\$”, “%”, “^”, “&”, “*”, “(”, “)”, “_”, “-”, “+”, “=”, “:”, ““”, “””, “<”, “>”, “koma”, “titik”, “?”, “/”, “\”, “#”, dan “|”. Penelitian ini tidak membuang simbol “@” karena simbol tersebut digunakan pada proses penanganan *mention*. Komentar yang mengandung simbol dan angka seperti “*Alhamdulillah sudah punya \U0001f618*” setelah diproses maka menjadi “*alhamdulillah sudah punya*”.
- b. *Lowercase folding*
Lowercase folding adalah mengubah semua huruf menjadi huruf kecil agar term benar-benar sama (*lowercase*). Komentar “*Bahan lembut nyaman*” menjadi “*bahan lembut nyaman*”, huruf “*B*” kapital diubah menjadi “*b*” huruf kecil.
- c. Penanganan *mention*
Mention adalah memanggil akun pengguna lain dalam sebuah komentar. Penulisan *mention* menggunakan simbol @namaakun. Karena nama akun yang di *mention* berbeda beda pada setiap komentar, maka diseragamkan dengan @mention agar menjadi 1 term yang memiliki makna yang sama. Komentar yang mengandung *mention* seperti “*Mau dong @adamwicaksono55 @adamwahyudi_11*” setelah proses penanganan *mention* menjadi “*mau dong @mention @mention*”
- d. *Word Normalizer*
Word Normalizer digunakan untuk memperbaiki kata-kata dalam kalimat sehingga menghasilkan kalimat yang sangat baik dan benar sesuai dengan aturan tata bahasa. Peningkatan ini diperlukan untuk memudahkan pembaca memahami makna kalimat [12]. Komentar “*Harga sdh termasuk ongkir kah*” maka setelah proses *Word Normalizer* menjadi “*harga sudah termasuk ongkir kah*”. Kata “*sdh*” diubah menjadi “*sudah*” sehingga lebih mudah dipahami. Istilah asing tidak dapat dinormalkan dan tetap digunakan pada penelitian ini, seperti kata “*ready*”, “*size*”, “*store*”, dll.
- e. *Tokenizer*
 Fungsi fitur ini adalah untuk membagi teks *input* menjadi *array token*, karakter yang digunakan dalam Bahasa Indonesia adalah alfabet dan setiap *token* dipisahkan oleh spasi[13]. Hasil dari *word normalizer* “*kalau untuk daerah sumbar toko ada dimana ya*” kemudian diubah menjadi “*kalau*”, “*untuk*”, “*daerah*”, “*sumbar*”, “*toko*”, “*ada*”, “*dimana*” dan “*ya*”.

Algoritma TF-IDF digunakan untuk melakukan perhitungan bobot komentar dan mengklasifikasikannya ke dalam 2 kelas (komentar potensial dan komentar tidak potensial). Pembobotan TF-IDF umumnya digunakan dalam penambangan teks dan pencarian informasi untuk mengevaluasi pentingnya istilah linguistik (umumnya *unigram* atau *bigram*) dalam korpus yang diteliti [14]. Perhitungan TF-IDF menggunakan persamaan 1.

$$tf_{t,d} = \frac{f_{t,d}}{n_d} \tag{1}$$

$$idf_t = \log \frac{N}{df_t}$$

$$W_{t,d} = tf_{t,d} \cdot idf_{t,d}$$

Keterangan:

$f_{t,d}$ = frekuensi term dalam dokumen d

df_t = frekuensi dokumen term t , yaitu jumlah di mana dokumen term t muncul

Analisis hasil dengan melakukan pengujian pada sistem menggunakan *confussion matrix* seperti pada Tabel 1 [5].

Tabel 1. *Confussion Matrix*

Kelas sebenarnya	Kelas Hasil Prediksi		
		Positif	Negatif
	Positif	True Positif (TP)	False Negatif (FN)
	Negatif	False Positif (FP)	True Negatif (TN)

Dimana,

- True Positif* = jumlah data dengan kelas positif dan hasil prediksi benar
- True Negatif* = jumlah data dengan kelas negatif dan hasil prediksi benar
- False Positif* = jumlah data dengan kelas positif dan hasil prediksi salah
- False Negatif* = jumlah data dengan kelas negatif dan hasil prediksi salah

Dari *confussion matrix* kemudian dilakukan perhitungan untuk mendapatkan nilai akurasi. Akurasi adalah salah satu langkah yang paling umum digunakan untuk kinerja klasifikasi, dan didefinisikan sebagai rasio antara sampel yang diklasifikasikan dengan benar dengan jumlah total sampel, akurasi dihitung menggunakan persamaan 2. Kemudian menghitung nilai presisi, presisi mewakili proporsi sampel positif yang diklasifikasikan dengan benar terhadap jumlah total sampel yang di prediksi positif sebagaimana ditunjukkan pada persamaan 3. *Recall* mewakili sampel positif yang diklasifikasikan dengan benar ke total jumlah sampel positif, dihitung menggunakan persamaan 4 [15].

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \tag{2}$$

$$Presisi = \frac{TP}{FP+TP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Pada dasarnya, nilai presisi dan *recall* adalah antara 0-1. Oleh karena itu, sistem pengambilan informasi yang baik diharapkan dapat memberikan nilai presisi dan penarikan mendekati 1 [11].

3. Hasil dan Pembahasan

3.1. Daftar akun toko Instagram

Akun toko yang digunakan berdomisili di Indonesia. Pencarian akun ini dilakukan pada awal bulan Februari 2020, menggunakan *hashtags* nama kategori produk. . Daftar akun toko ditunjukkan pada Tabel 2.

Tabel 2. Daftar akun toko di Instagram

No	Hashtags	Username	Kategori Produk	Followers	Komentar potensial	Komentar tidak potensial
1	#sepatuwanita	@ezshoesofficia	Sepatu wanita	407.000	24	5
2	#bajubayi	@velvetjunior	Pakaian bayi	460.000	20	8
3	#jam tangan	@gshockindo	Jam tangan	85.000	21	9
4	#sprei	@rizqirafli_sprei	Sprei dan beski	123.000	30	0
5	#hijabvoal	@nuqami	Hijab Voal	285.000	1	33
6	#kacamata	@yourglasses_id	Kaca mata	555.000	27	2
7	#cincinkawin	@qaylajewelry	Cincin kawin	390.000	24	14
8	#skincarekorea	@somebymi.official_id	Skin care korea	51.500	21	7
9	#handuk	@howelandco	Handuk	175.000	25	5
10	#sepatupria	@dobel_a	Sepatu pria	60.200	22	7

Pada Tabel 2, akun Instagram @rizqirafli_sprei memiliki jumlah komentar potensial terbanyak dan tidak memiliki komentar tidak potensial. Hasil ini berdasarkan pelabelan komentar dari 8 responden. Bisa disimpulkan bahwa pelanggan yang berkomentar pada akun toko @rizqirafli_sprei berpotensi dan memiliki niat untuk membeli produk. Akun toko dengan jumlah komentar tidak potensial terbanyak adalah @nuqami. Sebanyak 34 komentar hanya 1 yang memiliki label potensial, hal ini menjadi temuan bahwa

komentar banyak pada akun toko tidak memberikan jaminan bahwa pelanggan yang berkomentar memiliki niat untuk membeli.

3.2. Analisis Komentar

Data komentar yang telah dilabeli secara manual oleh responden kemudian dilakukan pra pemrosesan terlebih dahulu agar lebih mudah dipahami oleh sistem. Karena algoritma klasifikasi kami menggunakan TF-IDF di mana makna kata yang sama namun penulisan yang berbeda maka dianggap *term* yang berbeda, contoh variasi penulisan kata yang memiliki makna sama ditunjukkan pada Tabel 3.

Tabel 3. Variasi penulisan kata pada komentar yang memiliki makna sama

No	Kata pada komentar	Word Normalizer	Jumlah komentar
1	tidak	tidak	2
2	tdk	tidak	1
3	gak	tidak	17
4	g	tidak	3
5	ga	tidak	12
6	gk	tidak	3
7	berapa	berapa	20
8	brpa	berapa	5
9	brp	berapa	20
10	berapaan	berapa	3

Kata-kata pada tabel di atas adalah kata-kata yang memiliki variasi terbanyak yang digunakan pengguna pada fitur komentar. Kata “gak” adalah kata yang paling sering digunakan oleh pengguna dalam mengungkapkan kata “tidak” dan ditemukan pada 17 komentar, selain itu kata “brp” yang memiliki makna sebenarnya “berapa” ditemukan pada 20 komentar. Penggunaan *Word Normalizer* sangat dibutuhkan pada kasus ini karena banyaknya variasi penulisan kata dengan makna sama yang digunakan oleh pengguna Instagram.

Responden memberi label pada 300 komentar, kemudian label yang terbanyak dijadikan label pada *dataset*. *Dataset* komentar terdiri dari 205 dengan label komentar potensial, 89 komentar tidak potensial dan 6 komentar yang memiliki jumlah label sebanding. Label sebanding di mana 4 responden mengatakan bahwa komentar tersebut potensial dan 4 responden lainnya mengatakan bahwa komentar tersebut tidak potensial. Oleh karena itu komentar tersebut tidak digunakan pada perhitungan TF-IDF sehingga *dataset* menjadi 294 komentar dengan komposisi 194 data latih dan 100 data uji. Pada label komentar potensial sebanyak 43 komentar yang mutlak potensial, artinya seluruh responden sepakat mengatakan bahwa komentar tersebut potensial. Sedangkan pada label komentar tidak potensial memiliki 41 komentar yang mutlak tidak potensial. Hasil label komentar oleh responden ditunjukkan pada Tabel 4 dan Tabel 5.

Tabel 4. Hasil label komentar potensial oleh responden

No	Komentar Potensial	
	Sebelum pra pemrosesan	Setelah pra pemrosesan
1	Yang hitam size 38 masih kak	yang hitam size masih kak
2	Apakah barang ini masih tersedia?	apakah barang ini masih tersedia
3	di store purwodadi ready g kak?	di store purwodadi ready tidak kak
4	Sepasang harga berapa Om..	sepasang harga berapa om
5	ini sudah ada di store Tangerang belum ya?	ini sudah ada di store tangerang belum ya
6	yang putih no 37 ada gak kak?	yang putih no ada tidak kak
7	ada warna hitam kah?	ada warna hitam kah
8	Kl order kemana?	Kalau order kemana
9	Cara pesen gmn yaa	cara pesan bagaimana ya
10	Kak motif ini ada warna lain gak ?	kak motif ini ada warna lain tidak

Komentar pada Tabel 4. Adalah contoh dari beberapa komentar potensial oleh responden. Salah satu indikator minat beli adalah ketertarikan dalam mencari informasi produk[6], hal ini sesuai dengan hasil komentar potensial oleh responden di mana indikator ini diwakili oleh kata “masih”, “apakah”, “ready”, “kemana”, “bagaimana”, “ada”, “tidak”, dan “harga”.

Tabel 5. Hasil label komentar tidak potensial oleh responden

No	Komentar tidak potensial	
	Sebelum pra pemrosesan	Setelah pra pemrosesan
1	@ltssa	@mention
2	Mau dong @adamwicaksono55 @adamwahyudi_11	mau dong @mention @mention
3	Alhamdulillah sudah punya \U0001f618	alhamdulillah sudah punya
4	Nyaman	Nyaman
5	Masyaallah\U0001f60d	Masyaallah
6	Aman dan nyaman	aman dan nyaman
7	racun bat	racun bat
8	bagus banget	bagus banget
9	Bismillah, aku mau chesyy. Hijab yang eksklusif banget dan selalu kekinian \U0001f60d\U0001f60d\U0001f60d @rika1112003 @reario18 @utz_sriutam	bismillah aku mau chesyy hijab yang eksklusif banget dan selalu kekinian @mention @mention @mention
10	masyaallah mau banget	masyaallah cantik banget

Pada komentar tidak potensial, aktivitas *mention* lebih banyak ditemukan seperti pada nomor 1, 2, 3, dan 9. Nama *mention* yang diberikan oleh pengguna umumnya berbeda-beda sehingga diseragamkan dengan @mention. Penanganan *mention* ini juga membantu TF-IDF dalam mengenali aktivitas *mention* sebagai *term* yang sama. Pada komentar tidak potensial ditemukan kata-kata yang menunjukkan keinginan untuk memiliki seperti kata “*mau*” pada nomor 3, 9, dan 10. Rasa ingin memiliki juga adalah indikator minat beli, namun pada penelitian ini seluruh responden menganggap bahwa komentar tersebut tidak potensial. Hal ini perlu diteliti lebih lanjut. Selain itu pada komentar tidak potensial, tidak ditemukan kata-kata yang menunjukkan keingintahuan pada produk yang di-*posting*. Hal ini dibuktikan tidak adanya kata-kata tanya terkait produk seperti “*bagaimana*”, “*dimana*” dan “*apakah*”.

3.2.1. Penghitungan menggunakan TF-IDF

TF-IDF adalah statistik numerik yang dimaksudkan untuk mencerminkan betapa pentingnya sebuah kata dalam dokumen atau *corpus*. Hasil perhitungan TF-IDF pada 300 *dataset* ditunjukkan pada Tabel 6.

Tabel 6. Hasil perhitungan TF IDF

No	Term	DF	IDF	No	Term	DF	IDF
1	kak	58	0,538	14	dan	15	1,125
2	tidak	45	0,648	15	ready	15	1,125
3	ada	45	0,648	16	bagaimana	14	1,155
4	ya	33	0,783	17	bisa	14	1,155
5	ini	33	0,783	18	cantik	13	1,187
6	@mention	32	0,796	19	order	13	1,222
7	berapa	31	0,81	20	saya	12	1,222
8	yang	25	0,903	21	masih	12	1,222
9	mau	23	0,903	22	untuk	12	1,222
10	kalau	21	0,939	23	lensa	12	1,26
11	harga	17	0,979	24	sama	11	1,26
12	di	15	1,071	25	store	11	
13	minimal	15	1,071				

Pada Tabel 5 hanya ditampilkan *term* yang memiliki frekuensi kemunculan minimal 15 dokumen. *Term* yang muncul lebih dari satu kali pada satu dokumen yang sama tetap dihitung 1. Nilai DF tinggi artinya TF-IDF menjadi rendah [16]. Hasil pengujian 100 komentar ditunjukkan pada Tabel 7.

Tabel 7. Hasil pengujian *Confussion Matrix* 100 data uji pada Algoritma TF-IDF

Klasifikasi	Kelas prediksi		
	Potensial	Tidak potensial	
Kelas sebenarnya	Potensial	53	3
	Tidak potensial	17	27

Penelitian sebelumnya [10] menggunakan algoritma *Naïve Bayes* dalam mengklasifikasikan komentar pada Instagram memperoleh akurasi 76,5%. Sedangkan menggunakan TF-IDF pada penelitian memperoleh akurasi yang lebih baik yaitu 80%. Perbedaan akurasi mungkin disebabkan karena pada penelitian sebelumnya tidak menggunakan *word normalizer* sehingga *typo* penulisan kata yang disingkat

tidak dapat ditangani dengan baik. Seperti yang diketahui umumnya pada komentar Instagram banyak menggunakan Bahasa yang tidak baku, *typo*, dan disingkat.

Pada penelitian sebelumnya [11], data yang digunakan adalah artikel atau berita yang mana sudah memiliki Bahasa yang baik, berbeda dengan penelitian ini menggunakan data komentar. Penelitian sebelumnya nilai presisi 0,5 sedangkan pada penelitian ini sistem cukup baik dalam memprediksi komentar tidak potensial dengan nilai presisi 0,76, dari 70 komentar potensial yang diujikan sistem mampu memprediksi 53 komentar dengan benar dan nilai *recall* adalah 0,94.

Peneliti menggambarkan frekuensi kata yang sering digunakan oleh pengguna pada toko di Instagram dalam bentuk *wordcloud* ditunjukkan pada Gambar 3 .



Gambar 3a. *Wordcloud* komentar potensial



Gambar 3b. *Wordcloud* komentar tidak potensial

Gambar 3. *Wordcloud* komentar berdasarkan hasil akurasi TF-IDF.

Hasil komentar yang diklasifikasikan dengan benar oleh TF-IDF kemudian dirangkum dalam bentuk *wordcloud* menggunakan aplikasi WordItOut. Pada Gambar 3a. kata yang memiliki frekuensi kemunculan lebih dari 10 kali adalah kata “*kak*” 23 kali, “*berapa*” 18 kali, “*ini*” 17 kali, “*tidak*” dan “*ada*” sebanyak 13 kali. Kata “*kak*” sering digunakan oleh pengguna dalam bertanya sebuah produk seperti ditunjukkan pada Tabel 2, di mana kata tersebut digunakan pada baris data nomor 1, 3, 6, dan 10. Sedangkan pada Gambar 3b. untuk kata dominan pada komentar tidak potensial adalah kata “*mention*”, “*chessy*”, dan “*mau*”. Analisis respons pengguna dan pengguna potensial pada komentar *posting-an* produk yang diperoleh penjual ditunjukkan pada Tabel 8.

Tabel 8. Hasil analisis komentar pada *posting-an* produk.

Url post	B73JTvhhXFm
Username	@velvetjunior
Kategori	Pakaian bayi
Caption	AMAN & SEHAT adalah hal utama yang harus menjadi pertimbangan kita dalam memilih pakaian untuk buah hati tercinta, selain karena kelembutan serta motifnya yang lucu dan menarik.\nAman & sehat dari apa ya? Adakah Bunda yang tahu?\nYuk, share lagi di kolom komentar ya... #velvetjunior #trustvelvetjunior #bajubayi #bajuanak
Jumlah komentar	12
Jumlah pengguna yang berkomentar	12
Respon pengguna	dimana, daerah, pesan, harga, order, toko, kak
Pengguna potensial	@djingga_02, @nurulhaqcollection, @fitriendra, @icakhoeruanisa, @nrita, @anggiupra, @lialia1561, @risellanasution, @elvibahderlys, @isma2309

Tabel 8 adalah salah satu hasil analisis komentar pada *posting-an* produk pakaian bayi, komentar sudah dilakukan pra pemrosesan Berdasarkan hasil klasifikasi diperoleh 10 pelanggan yang potensial atau berminat untuk membeli. Kata pada label respons pengguna yang ditampilkan adalah yang ditulis lebih dari 1 kali. Kata terbanyak yang ditulis oleh pengguna pada komentar *posting* produk tersebut adalah kata “*dimana*” yang muncul pada 4 komentar. Kata “*dimana*” digunakan pengguna untuk menanyakan lokasi *store* dan asal alamat pengiriman. Sebanyak 3 komentar menggunakan kata “*pesan*”, kata tersebut digunakan oleh pengguna untuk menunjukkan minat beli dan bertanya mengenai prosedur pemesanan.

4. Kesimpulan

Komentar pada toko Instagram dibutuhkan oleh pelanggan dalam mencari informasi tentang produk dan *review* dari pengguna lain. Sedangkan untuk penjual komentar dapat digunakan untuk melihat respons tentang produk yang dijual dan mendapatkan pelanggan yang potensial. Namun banyaknya komentar yang ada pada *posting* toko Instagram tidak menjamin bahwa pengguna yang berkomentar berniat membeli, dibuktikan dengan 294 komentar, 27% di antaranya adalah komentar tidak potensial.

Pendekatan NLP untuk memperbaiki bahasa yang tidak baku dan singkatan pada komentar Instagram sangat diperlukan, seperti kata "*tidak*" memiliki 6 variasi penulisan yaitu "*tidak*", "*tdk*", "*gak*", "*ga*", "*g*", dan "*gk*". Fitur NLP yang dapat memperbaiki kata tersebut adalah *Word Normalizer* sehingga sistem dapat mengenali variasi tersebut dan menjadikannya 1 *term* yang sama.

Algoritma TF-IDF dapat mengklasifikasikan dengan baik komentar pada Instagram ke dalam dua kelas yaitu komentar potensial dan komentar tidak potensial, dengan nilai akurasi sebesar 80%, presisi 0,76 dan *recall* 0,94. Penelitian ini menemukan bahwa kata-kata yang menunjukkan minat beli seseorang pada Instagram adalah kata "*berapa*", "*kak*", "*ada*", dan "*tidak*". Sedangkan kata yang paling dominan dalam menunjukkan komentar tidak potensial adalah kata "*mention*" yang mewakili aktivitas *mention*.

Hasil dari penelitian ini dapat diterapkan untuk sistem analisis pemasaran di mana dari komentar potensial penjual dapat mengumpulkan pengguna Instagram yang tepat untuk produknya, sehingga aktivitas pemasaran melalui Instagram lebih tepat sasaran. Selain itu komentar potensial juga dapat dimanfaatkan untuk deteksi produk dan analisis kompetitor.

Daftar Pustaka

- [1] M. Mohammadi, M. Dawodi, W. Tomohisa, and N. Ahmadi, "Comparative study of supervised learning algorithms for student performance prediction," *1st Int. Conf. Artif. Intell. Inf. Commun. ICAIIC 2019*, pp. 124–127, 2019.
- [2] W. K. Pertiwi, "Riset Ungkap Pola Pemakaian Medsos Orang Indonesia," *Kompas*, 2018. [Online]. Available: <https://tekno.kompas.com/read/2018/03/01/10340027/riset-ungkap-pola-pemakaian-medsos-orang-indonesia>. [Accessed: 02-Feb-2020].
- [3] D. Setyowati, "Instagram Klaim 70% Pengguna Akun Bisnisnya Serap Tenaga Kerja," *Kata Data*, 2019. [Online]. Available: <https://katadata.co.id/berita/2019/03/05/instagram-klaim-70-pengguna-akun-bisnisnya-serap-tenaga-kerja>. [Accessed: 02-Feb-2020].
- [4] O. Gibreel, D. A. Alotaibi, and J. Altmann, "Social commerce development in emerging markets," *Electron. Commer. Res. Appl.*, no. December, 2017.
- [5] A. R. C, Y. Lukito, P. T. Informatika, F. T. Informasi, U. Kristen, and D. Wacana, "Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes," no. August, 2017.
- [6] S. Adinda, "PENGARUH MEDIA SOSIAL INSTAGRAM @exploremalang TERHADAP MINAT BERKUNJUNG FOLLOWERS KE SUATU DESTINASI (SURVEI PADA FOLLOWERS @exploremalang)," *J. Adm. Bisnis*, vol. 72, no. 1, 2019.
- [7] D. I. Anjaskara, "Pengaruh Sikap pada Media Sosial Instagram terhadap Minat Beli Produk Kecantikan melalui Instagram," 2016.
- [8] A. S. Al-adwan and H. Kokash, "The Driving Forces of Facebook Social Commerce," *J. Theor. Appl. Electron. Commer. Res.*, vol. 14, no. 2, pp. 15–32, 2019.
- [9] S. M. Din, R. Ramli, and A. A. Bakar, "A Review on Trust Factors affecting Purchase Intention on Instagram," in *6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, 2017, pp. 49–53.
- [10] A. Fauzi, M. F. Akbar, and Y. F. A. Asmawan, "Sentimen Analisis Berinternet Pada Media Sosial dengan Menggunakan Algoritma Bayes," *J. Inform.*, vol. 6, no. 1, pp. 77–83, 2019.
- [11] A. N. Khusna and I. Agustina, "Implementation of Information Retrieval Using Tf-Idf Weighting Method On Detik.Com's Website," in *2018 12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 2018, pp. 1–4.
- [12] A. N. Rohman, E. Utami, and S. Raharjo, "Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing," *Eksplora Inform.*, vol. 9, no. 1, pp. 70–76, 2019.
- [13] A. Purwarianti, A. Andhika, A. F. Wicaksono, I. Afif, and F. Ferdian, "InaNLP: Indonesia natural language processing toolkit, case study: Complaint tweet classification," *4th IGNITE Conf. 2016 Int. Conf. Adv. Informatics Concepts, Theory Appl. ICAICTA 2016*, pp. 5–9, 2016.
- [14] I. Yahav, O. Shehory, and D. Schwartz, "Comments Mining With TF-IDF: The Inherent Bias and Its Removal," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 3, pp. 437–450, 2019.

- [15] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, 2018.
- [16] O. W. Purbo, *Text Mining Analisis MedSos, Kekuatan Brand & Intelejen di Internet*, 1st ed. Yogyakarta, Indonesia: ANDI, 2019.