

PEMANFAATAN ANALISIS SENTIMEN DARI ULASAN PRODUK DI YOUTUBE UNTUK PENGEMBANGAN PRODUK BARU

Ricky Paian Limbong¹, Ronsen Purba², Muhammad Fermi Pasha³

Universitas Mikroskil, Medan, Indonesia^{1,2,3}

Email: ricky.limbong@mikroskil.ac.id¹, ronsen@mikroskil.ac.id²,
muhammad.pasha@mikroskil.ac.id³

Abstrak

Pengembangan produk yang sukses memerlukan pemahaman tentang kebutuhan dan preferensi pelanggan. Analisis sentimen telah muncul sebagai alat yang dapat mengumpulkan pendapat dari pelanggan dalam mengembangkan yang lebih baik. Penelitian ini bertujuan untuk mengeksplorasi pemanfaatan analisis sentimen dari ulasan produk di YouTube dalam rangka pengembangan produk baru. Dengan menganalisis konten yang dibuat oleh pengguna, penelitian ini bertujuan untuk menghasilkan informasi berupa prioritas fitur produk. Metode penelitian meliputi pengumpulan dan prapemrosesan data ulasan produk dari platform YouTube, dengan menerapkan teknik pemrosesan teks seperti *case folding*, penghilangan kata yang tidak relevan, tokenisasi, dan *stemming*. Analisis sentimen dilakukan menggunakan metode *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen yang diekspresikan dalam ulasan tersebut. Model yang telah dilatih kemudian digunakan untuk memprediksi dan memberi label sentimen pada ulasan produk baru. Temuan penelitian ini menunjukkan bahwa analisis sentimen dapat membantu proses pengembangan produk baru dengan memperhatikan prioritas fitur produk yang memiliki kekurangan. Pendekatan ini memungkinkan perusahaan untuk memahami kebutuhan pelanggan, membuat keputusan yang tepat dalam memberikan fokus untuk peningkatan fitur produk untuk perilsan selanjutnya. Integrasi analisis sentimen dalam proses pengembangan produk baru dapat memanfaatkan opini konsumen untuk merilis produk yang lebih baik.

Kata kunci: Analisis Sentimen; Youtube; Pengembangan Produk Baru

Abstract

Successful product development requires an understanding of customer needs and preferences. Sentiment analysis has emerged as a tool capable of gathering customer opinions to enhance product development. This research aims to explore sentiment analysis from product reviews on YouTube for new product development. By analyzing user-generated content, this study seeks to generate insights in the form of product feature priorities. The research methodology involves collecting and preprocessing product review data from the YouTube platform, employing text processing techniques such as case folding, removal of irrelevant words, tokenization, and stemming. Sentiment analysis is conducted using the Support Vector Machine (SVM) method to classify sentiments expressed in the reviews. The trained model is then employed to predict and label sentiments in new product reviews. Findings of this study indicate that sentiment analysis can aid the new product development process by highlighting feature priorities that require attention. This approach enables companies to comprehend customer needs and make informed decisions in focusing

on feature enhancements for subsequent releases. Integrating sentiment analysis into the new product development process can leverage consumer opinions to launch improved products.

Keywords: *Sentiment Analysis; Youtube; New Product Development*

Pendahuluan

Analisis sentimen merupakan proses otomatis untuk mengenali dan menangkap informasi mengenai opini yang diungkapkan pengguna dalam sebuah teks (Dang dkk., 2020). Proses ini dapat mengekstraksi informasi subjektif dari data teks yang besar dan tidak terstruktur, seperti ulasan produk di platform *e-commerce* dan unggahan di media sosial. Dari sudut pandang merek, analisis sentimen dapat memberikan wawasan yang berharga tentang opini, preferensi, dan kebutuhan pelanggan, yang dapat digunakan untuk berbagai tujuan, seperti menangani layanan pelanggan (Kodhai dkk., 2020), meningkatkan sistem rekomendasi (Kauffmann dkk., 2019), dan identifikasi kebutuhan pelanggan dari ulasan produk (Ireland & Liu, 2018). Berdasarkan laporan dari *datareportal.com*, mencari produk yang ingin dibeli dan berbagi pendapat merupakan salah satu alasan pengguna internet menggunakan media sosial. Dapat disimpulkan bahwa platform media sosial yang banyak digunakan untuk mengulas produk adalah YouTube, yang telah menjadi saluran penting bagi konsumen untuk berinteraksi dengan merek dan membuat keputusan pembelian (Huang dkk., 2022). Namun, analisis sentimen dari ulasan produk di YouTube memiliki tantangan dan peluang tersendiri, seperti tantangan dalam mengolah data teks yang bervariasi, tidak terstruktur, dan mengandung bahasa informal, *slang*, atau ironi (Pradhan, 2021). Kehadiran media sosial YouTube memainkan peran penting dalam membentuk persepsi pelanggan dan mempengaruhi perilaku pembelian di industri elektronik terutama industri *smartphone* (Frick & Kaimann, 2017).

Untuk memenangkan persaingan dan menarik perhatian konsumen, perusahaan-perusahaan di industri elektronik ini harus terus berinovasi dan memperkenalkan produk-produk baru yang sesuai dengan kebutuhan dan preferensi konsumen (Frick & Kaimann, 2017). Salah satu cara untuk berinovasi adalah dengan melibatkan komunikasi umpan balik konsumen dalam proses *new product development* (NPD) (Hamid & ABBASI, 2020). NPD adalah proses yang dinamis dan inovatif yang cepat dan dapat diandalkan untuk membantu keberhasilan terhadap produk baru yang akan diluncurkan, proses ini memungkinkan perusahaan untuk memberikan pembaruan terhadap produk yang akan dirilis berdasarkan produk sebelumnya (Thirumagal Vijaya, 2014). Keterlibatan konsumen dalam proses NPD terjadi pada fase pertama yaitu *Idea Generation* (Pembangkitan Ide) yang memberikan dampak yang positif untuk keberhasilan NPD (Thirumagal Vijaya, 2014). Berbagai metode dapat digunakan untuk mengumpulkan masukan ataupun komunikasi umpan balik dari konsumen, salah satu metode yang paling banyak digunakan adalah *text mining* (Latuny dkk., 2021).

Penggunaan analisis sentimen untuk NPD menimbulkan beberapa tantangan yang harus diatasi oleh para peneliti terdahulu, seperti mengeksplorasi dan menentukan cara yang efektif untuk menerapkan analisis sentimen dalam NPD. Peneliti bertujuan untuk menyelidiki bagaimana umpan balik pelanggan dan kecenderungan emosional dapat digunakan untuk meningkatkan desain dan pengembangan produk. Namun, setiap peneliti memiliki solusi yang berbeda-beda untuk mencapai hasil yang optimal. Sutrilastyo & Diar Astanti menelusuri penggunaan analisis sentimen berbasis Lexicon

untuk mengetahui apa yang menjadi ekspektasi dan permintaan pelanggan untuk desain *underbody* mobil, hasilnya adalah perusahaan dapat memeriksa ulang apakah definisi kriteria pelanggan selaras dengan kriteria perusahaan (Sutrilastyo & Diar Astanti, 2021). Giannakis mengajukan *fuzzy comprehensive evaluation* yang digunakan pada *sentiment analysis* untuk mengukur kecenderungan emosional kualitatif dan untuk menghitung nilai emosional detail terhadap produk (Giannakis dkk., 2022). Rathore & Ilavarasan mengembangkan pendekatan terhadap perubahan emosi untuk produk sebelum rilis dan setelah rilis. Hasil perbandingannya menunjukkan pergeseran niat perilaku pengguna terhadap produk baru dengan mempertimbangkan kinerja pasarnya (Rathore & Ilavarasan, 2020).

Penelitian ini bertujuan untuk menghasilkan informasi yang dapat digunakan dalam proses pengembangan produk baru (NPD) pada tahap *Idea Generation*. Informasi ini akan diperoleh melalui analisis sentimen terhadap kata kunci produk secara langsung. Bentuk hasil pemanfaatan analisis sentimen pada penelitian ini adalah prioritas perbaikan produk dengan bantuan visualisasi data yang akan menampilkan urutan fitur produk yang paling banyak dikomentari, fitur produk yang memiliki jumlah sentimen negatif yang lebih banyak dari sentimen positif, dan urutan jumlah sentimen fitur produk dari yang terbesar hingga terkecil.

Metode Penelitian

Sumber data umpan balik konsumen bersumber dari komentar YouTube dalam konten ulasan produk. Analisis Sentimen terhadap komentar YouTube dapat memberikan kesimpulan informasi bahwa apakah produk yang diulas menghasilkan umpan balik yang positif, atau negatif, atau netral. Hasil informasi ini dapat memberikan masukan bagi *brand* terhadap produknya, namun informasi ini masih bersifat umum dan belum memberikan gambaran yang detail tentang aspek-aspek spesifik produk yang menjadi sorotan atau kritikan konsumen. Dalam hal ini, diperlukan informasi yang lebih spesifik mengenai sentimen langsung terhadap bagian-bagian produk yang diulas sehingga dapat menghasilkan masukan yang dapat digunakan pada tahap *Idea Generation* dalam proses NPD.

Salah satu pendekatan yang lebih efisien dan efektif dalam mengumpulkan informasi umpan balik konsumen terhadap produk baru adalah analisis sentimen dengan memanfaatkan algoritma *machine learning*, *Support Vector Machine* (SVM), yang dapat mengklasifikasikan sentimen konsumen dengan akurasi yang tinggi (Iskandar & Nataliani, 2021). Dengan memanfaatkan algoritma klasifikasi ini, dapat dilakukan pelabelan sentimen terhadap komentar YouTube dan dilanjutkan ke penerapan hasil sentimen untuk pengembangan produk baru pada tahap *Idea Generation*.

Penelitian ini memiliki 4 proses utama untuk menyelesaikan permasalahan yang telah diuraikan yaitu data praproses, ekstraksi pembobotan kata kunci, analisis sentimen, dan penerapan hasil analisis sentimen untuk fase *idea generation*.

a. Data praproses

Data praproses diawali dengan pemindahan data dari kolom "Reply" ke "Comment" untuk menggabungkan data komentar dan balasannya. Kemudian, kolom "Comment" difilter dan dijadikan *Dataframe* baru. Atribut HTML dihapus untuk membersihkan data. Selanjutnya, proses *data cleaning* dilakukan dengan langkah-langkah berikut:

1. *Casefolding*: mengubah semua huruf menjadi huruf kecil untuk menstandarkan teks.
2. Menghapus kata-kata tidak penting: tanda baca, angka, *whitespace*, dan *mention*.
3. Tokenisasi: membagi teks menjadi unit-unit kecil (*token*) untuk memudahkan proses analisis.
4. *Stemming*: memotong akhiran atau awalan kata untuk mengembalikan kata ke bentuk dasarnya.
5. Persiapan kamus dan normalisasi kata: memperbaiki kata-kata *typo* dan mengubah *keyword* Bahasa Inggris ke Bahasa Indonesia. Tabel 1 menampilkan isi dari kamus yang merupakan perbaikan untuk proses normalisasi kata-kata *typo* menjadi kata normal.

Tabel 1. Kamus Keyword Fitur *Smartphone*

indeks	keyword_fitur_typo	formal
0	audioy	speaker
1	microphone	mikrofon
2	radionya	radio
3	gbrnya	foto
4	harganyagak	harga
5	softwharenya	software
6	mesinnya	prosesor
7	bloatware	software
8	changer	pengecasan
9	kembung	bengkak
10	cipsetnya	chipset
11	speed	kecepatan
12	bezzelnya	layar
13	mediaatek	prosesor
14	snpdrgon	prosesor
15	qualcomm	prosesor
16	browser	aplikasi
17	desaign	desain
18	socnya	prosesor
19	designa	desain
20	usb	usb
21	fto	foto
22	bklan	akan
23	cipset	chipset
24	jazz	bagus
25	kreen	keren
26	ngebut	kencang
27	diplay	layar
28	cpu	prosesor
29	ufs	penyimpanan
30	mediatex	prosesor
31	kamerax	kamera
32	dimencity	prosesor
33	battery	baterai
34	lemotnya	lambat
35	built	body

indeks	keyword_fitur_typo	formal
36	qualtynya	kualitas
37	ufsnya	penyimpanan
38	baretnya	baterai
39	apes	sial
40	processor	prosesor
41	desainnyaaa	desain
42	stabilisasi	stabilizer
43	gpu	kartu grafis
44	pemroses	prosesor
45	bloatwarenya	software
46	fotografi	kamera
47	spiker	speaker
48	sdragon	prosesor
49	touchsampling	layar
50	suaranyaaaaa	speaker
51	hargany	harga
52	stylenya	desain
53	memorynya	memori
54	mainboardnya	prosesor
55	notch	layar
56	uinya	UI
57	processornya	prosesor
58	memorinya	memori
59	mediateeek	prosesor
60	stabilisernya	stabilizer
61	batreinya	baterai
62	speker	speaker
63	procie	prosesor
64	quotqualcom	prosesor
65	snapdragonquot	prosesor
66	qualcom	prosesor
67	bentuknya	body
68	desingan	desain
69	mitektek	prosesor
70	dimesity	prosesor
71	dikamernya	kamera
72	hp	handphone
73	ampas	buruk
74	pnas	pnas
75	kecang	kencang
76	kelewatan	berlebihan
77	lag	lambat
78	ngeblur	buram

b. Ekstraksi pembobotan kata kunci

Setelah dilakukan preproses terhadap dataset, selanjutnya adalah melakukan *Term-weighting* atau Pembobotan kata. *Term-weighting* merupakan proses pemberian

bobot term pada dokumen. Pembobotan ini digunakan nantinya oleh algoritma *Machine Learning*, yaitu *Support Vector Machine* (SVM) untuk klasifikasi dokumen.

Kedua proses ini dijalankan secara bersamaan dengan menggunakan teknik dari library python *TfidfVectorizer*. *TfidfVectorizer* merupakan bagian dari proses ekstraksi dan pembobotan kata dalam analisis teks. *TfidfVectorizer* menggabungkan dua konsep utama dalam pemrosesan teks, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF), untuk menghasilkan representasi numerik dari kata-kata dalam teks. Proses ekstraksi kata-kata melibatkan tokenisasi teks menjadi unit-unit yang lebih kecil. *TfidfVectorizer* kemudian menghitung frekuensi kemunculan setiap kata (TF) di dalam setiap dokumen dan menghitung invers dari frekuensi dokumen yang mengandung kata tersebut (IDF).

c. Analisis sentimen

Tahap ini dilakukan dengan pembuatan model pembelajaran mesin dengan menggunakan library SVM dari Python, yaitu modul *svm* dari library *scikit-learn*. Dengan menggunakan *svm.SVC(kernel='linear')*, kemudian membuat objek SVM dengan kernel linear sebagai model yang akan dilatih menggunakan data latih untuk mempelajari pola dan korelasi antara fitur (teks) dan label sentimen yang sesuai.

Proses pelatihan model SVM dilakukan menggunakan data latih dari repositori Github yang dimiliki oleh github.com/IndoNLP/indonlu dari indobenchmark.com (Wilie dkk., t.t.). Data latih yang digunakan berjumlah sebanyak 11000 baris dengan 2 buah kolom, yaitu Teks dan Target. Isi dataset data latih IndoNLU dapat dilihat pada Gambar 3.10. Dengan menggunakan *fit(train_vectors, data_train['Target'])* untuk melatih model SVM, di mana *train_vectors* adalah representasi vektor dari teks data latih, dan *data_train['Target']* adalah label sentimen yang sesuai dengan data latih. Isi dataset data latih IndoNLU dapat dilihat pada Gambar 1.

	Teks	Target
0	meski masa kampanye sudah selesai , bukan bera...	neutral
1	tidak enak	negative
2	restoran ini menawarkan makanan sunda . kami m...	positive
3	lokasi di alun alun masakan padang ini cukup t...	positive
4	betapa bejad kader gerindra yang anggota dprd ...	negative

Gambar 1. Data Uji IndoNLU (Wilie dkk., t.t.)

Dengan menggunakan *predict(test_vectors)* untuk melakukan prediksi sentimen pada data uji, di mana *test_vectors* adalah representasi vektor dari teks data uji. Hasil prediksi sentimen disimpan dalam variabel *predictions*, yang berisi label sentimen yang diprediksi oleh model SVM untuk setiap data uji. Hasil pengujian algoritma SVM ini dapat dilihat pada Gambar 2.

	precision	recall	f1-score	support
positive	0.81	0.86	0.84	394
negative	0.79	0.70	0.74	131
netral	0.92	0.91	0.91	735
accuracy			0.87	1260
macro avg	0.84	0.82	0.83	1260
weighted avg	0.87	0.87	0.87	1260

Gambar 2. Hasil Pengujian Algoritma SVM

d. Penerapan hasil analisis sentimen untuk fase *idea generation*

Ini merupakan tahap dimana hasil dari analisis sentimen komentar akan digunakan untuk menghasilkan informasi yang dapat digunakan untuk *Idea Generation* dalam proses NPD. Tahap yang akan dilaksanakan dalam proses ini, yaitu identifikasi kebutuhan pelanggan dan penentuan prioritas kebutuhan konsumen.

Identifikasi kebutuhan pelanggan

Identifikasi kebutuhan pelanggan dilakukan dengan mencari dan menyaring data komentar yang dimana secara langsung bagian produknya dikomentari. Untuk itu kata kunci fitur produk *smartphone* dipersiapkan untuk proses penyaringan nanti, daftar kata kunci dan kategorinya dapat dilihat pada Tabel 2. Penetapan kata kunci ini diambil dari situs web *gsmarena.com* yang menyediakan informasi, ulasan, spesifikasi, dan berita tentang ponsel, smartphone, tablet, dan perangkat seluler lainnya. Kemudian hasil penyaringannya akan dihitung jumlah sentimennya untuk masing-masing *keyword*, dan kemudian hasil perhitungan dapat dipresentasikan sebagai informasi untuk *Idea Generation* sebagai masukan untuk pengembangan produk baru.

Tabel 2. Daftar Kata Kunci untuk Penyaringan Data

No.	Smartphone Feature	Kategori
0	Tanggal rilis	Perilisan
1	Status	Perilisan
2	Body	Body
3	Dimensi	Body
4	Berat	Body
5	Desain	Body
6	Sim	Body
7	Layar	Layar
8	Tipe layar	Layar
9	Ukuran	Layar
10	Resolusi	Layar
11	Perlindungan	Layar
12	Platform	Platform
13	Software	Platform
14	Sistem operasi	Platform
15	Chipset	Platform
16	Cpu	Platform
17	Gpu	Platform
18	Memori	Memori
19	Ram	Memori

No.	Smartphone Feature	Kategori
20	Slot kartu	Memori
21	Memori internal	Memori
22	Kamera	Kamera
23	Kamera utama	Kamera
24	Fitur kamera utama	Kamera
25	Video kamera utama	Kamera
26	Kamera selfie	Kamera
27	Fitur kamera selfie	Kamera
28	Video kamera selfie	Kamera
29	Speaker	Suara
30	Jack 3.5mm	Suara
31	Wifi	Koneksi
32	Bluetooth	Koneksi
33	Gps	Koneksi
34	Nfc	Koneksi
35	Infrared	Komunikasi
36	Radio	Komunikasi
37	Usb	Komunikasi
38	Sensor	Sensor
39	Sidik jari	Sensor
40	Baterai	Baterai
41	Tipe baterai	Baterai
42	Pengisian baterai	Baterai
43	Warna	Lainnya
44	Model	Lainnya
45	Harga	Lainnya
46	Performa	Tes
47	Layar	Tes
48	Kamera	Tes
49	Pengeras suara	Tes
50	Daya tahan baterai	Tes

Kemudian, setelah dilakukan penyaringan data, maka jumlah baris akan berkurang sesuai dengan pencocokan baris komentar dengan kata kunci yang telah ditetapkan. Selanjutnya adalah membuat kolom baru dengan nama “Part of Product” untuk tempat ekstraksi kata kunci yang telah berhasil dicocokkan. Gambaran tabelnya dapat dilihat pada Tabel 3.

Tabel 3. Gambaran Tabel untuk Ekstraksi Kata Kunci ke Kolom Baru

Comment	Part Of Product	Sentiment

Tahap selanjutnya adalah membuat *dataframe* baru untuk membuat penghitungan sentimen negative, positif, dan netral terhadap masing-masing fitur kata kunci. Dalam hal ini digunakan fungsi *pivot table* dari *library* pandas. Kemudian pada saat proses pembuatan kolom tabel pivot akan membuat 3 kolom baru sesuai dengan nilai yang ada di kolom sentimen, yaitu kolom “Negative”, kolom “Positif”, dan kolom “Neutral”. Nilai

akan akan diisi pada masing-masing kolom adalah hitungan setiap sentimen untuk setiap nilai pada kolom “Part of Product”. Gambaran tabelnya dapat dilihat pada Tabel 4.

Tabel 4. Gambaran Tabel untuk Perhitungan Sentimen

Part of Product	Negatif	Positif	Netral

Penetapan Prioritas Kebutuhan Pelanggan

Pada tahap ini dilakukan pengurutan kata kunci berdasarkan jumlah frekuensi kata kunci dari komentar. Hasil pengurutan ditampilkan dengan bantuan *library* matplotlib. Pada hasil pengurutan ditampilkan juga ilustrasi warna untuk membedakan jumlah sentimen negatif, positif, dan netral. Untuk negatif diberi warna merah, dan positif diberi warna biru, dan untuk netral diberi warna hitam. Visualisasi data akan dibuat sebanyak tiga yang akan ditampilkan berdasarkan semua jumlah sentimen untuk masing-masing fitur yang dimana polaritas sentimen negatif, positif dan netral digabung, kemudian yang kedua menampilkan sentimen negatif dan positif untuk melihat fitur mana saja yang memiliki nilai negatif yang lebih banyak dari sentimen positif, dan yang terakhir mengurutkan urutan jumlah sentimen negatif untuk melihat fitur mana yang memiliki jumlah sentimen negatif yang terbesar dirutkan hingga ke terkecil.

Hasil dan Pembahasan

Hasil

Analisis Sentimen terhadap komentar Youtube melibatkan beberapa tahapan penting, termasuk Preproses Data, Ekstraksi dan Pembobotan Kata, serta Pemberian Label Sentimen. Preproses data mencakup penggabungan data, penghapusan kolom yang tidak relevan, penghapusan *tag* HTML, *casefolding*, penghapusan karakter dan kata yang tidak penting, tokenisasi, *stemming*, dan normalisasi kata. Proses ini bertujuan untuk mempersiapkan dataset agar siap untuk analisis lebih lanjut. Gambar 3 menunjukkan hasil akhir setelah dilakukan pembersihan data.

Tokenized_Comment	Stemmed_Comment
[ada, perbedaannya, ga, bang, sama, yg, redmi,...]	[ada, beda, ga, bang, sama, yg, redmi, note, pro]
[sebenarnya, mending, ini, apa, poco, ya, sara...]	[benar, mending, ini, apa, poco, ya, saran, do...]
[cocok, buat, ngegame, kah, ini, hp]	[cocok, buat, ngegame, kah, ini, hp]
[dari, dulu, selalu, jatuh, cinta, sama, xiaom...]	[dari, dulu, selalu, jatuh, cinta, sama, xiaom...]
[pengguna, redmi, redmi, note, sekarang, redmi...]	[guna, redmi, redmi, note, sekarang, redmi, no...]

Gambar 3. Hasil Akhir setelah Preproses Data

Ekstraksi dan Pembobotan Kata dilakukan dengan `TfidfVectorizer` untuk menghitung bobot TF-IDF setiap kata dalam setiap dokumen. Bobot TF-IDF mengindikasikan tingkat pentingnya sebuah kata dalam dokumen tersebut dan dalam keseluruhan korpus. Kata-kata yang muncul lebih sering dalam dokumen tertentu, namun jarang muncul di dokumen lain, akan memiliki bobot TF-IDF yang tinggi. Gambar 4 menunjukkan representasi *sparse* matriks TF-IDF yang dihasilkan oleh `TfidfVectorizer`. Setiap baris dalam output tersebut merepresentasikan sebuah dokumen atau teks, sedangkan setiap pasangan angka pada baris tersebut menunjukkan indeks kata dan bobot TF-IDF dari kata tersebut di dalam dokumen tersebut.

(0, 3188)	0.48773817702004385
(0, 3102)	0.4611613437533186
(0, 2665)	0.48773817702004385
(0, 385)	0.3309358695474968
(0, 327)	0.4494825988053424

Gambar 4 Hasil Output Setelah Vektorisasi

Sebagai contoh, pasangan (0, 3188) dengan nilai 0.48773817702004385 menandakan bahwa pada baris ke-0 (dokumen pertama) dan kolom ke-3188 (kata ke-3188), nilai TF-IDF adalah 0.48773817702004385. Hal ini menunjukkan bahwa kata ke-3188 memiliki tingkat penting yang cukup tinggi dalam dokumen pertama. Jika kita mengakses kata pada indeks ke-3188 menggunakan `vectorizer.get_feature_names_out()`, akan muncul kata "redmi", yang menunjukkan bahwa kata tersebut memiliki kepentingan yang signifikan dalam dokumen pertama. Dengan menggunakan representasi TF-IDF ini, maka dapat digunakan sebagai input untuk model SVM guna melakukan prediksi sentimen.

Pemberian Label Sentimen dilakukan dengan menerapkan model klasifikasi sentimen yang telah disimpan sebelumnya. Hasil dari pelabelan sentimen digunakan untuk proses pengembangan produk baru dalam fase *Idea Generation*. Hasil klasifikasi dan pelabelan sentimen dapat dilihat pada Gambar 5.

	Comment	Sentiment
0	ada perbedaannya ga bang sama yg redmi note pro	neutral
1	sebenarnya mending ini apa poco ya saran dongg	negative
2	cocok buat ngegame kah ini hp	positive
3	dari dulu selalu jatuh cinta sama xiaomi karna...	positive
4	pengguna redmi redmi note sekarang redmi note...	neutral

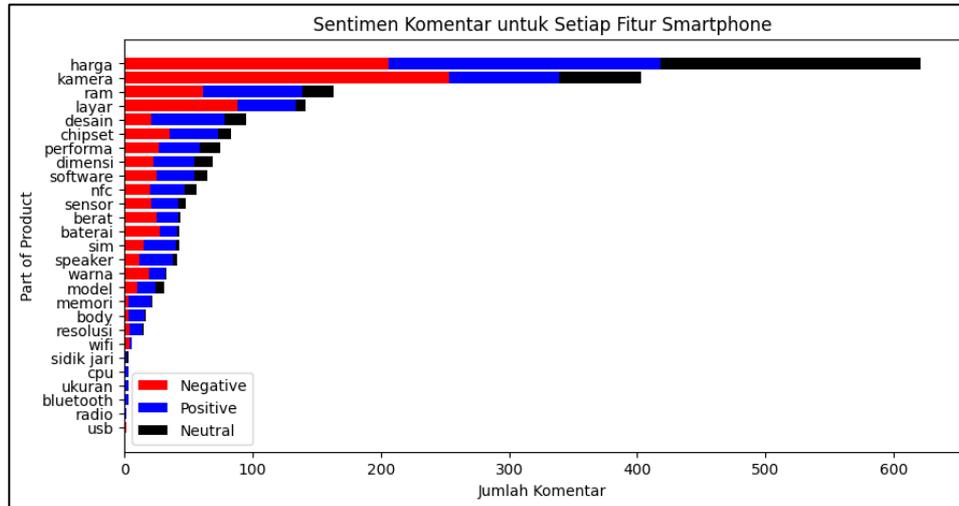
Gambar 5 Hasil Pelabelan Sentimen

Setelah pelabelan sentimen terhadap data komentar selesai dilakukan, dilakukan penyaringan data berdasarkan kumpulan kata kunci fitur produk. Data yang diinginkan adalah komentar yang merujuk pada topik yang sesuai dengan kamus fitur khusus produk smartphone. Selanjutnya, dibuat kolom baru yang diisi dengan kata kunci yang berhasil dicocokkan dengan kolom “Comment”. Setelah itu, dilakukan penjumlahan sentimen untuk masing-masing kata kunci di kolom Part of Product seperti yang terlihat pada Tabel 5.

Tabel 5. Penjumlahan Sentimen untuk masing-masing Kata Kunci

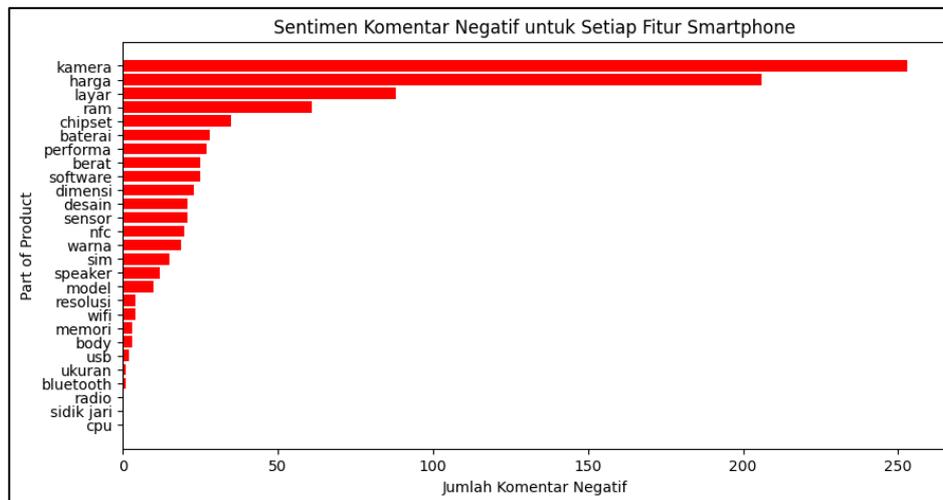
Part of Product	Negative	Positive	Neutral	Total Comments
baterai	28	13	2	43
berat	25	17	2	44
bluetooth	1	2	0	3
body	3	13	1	17
chipset	35	38	10	83
cpu	0	3	0	3
desain	21	57	17	95
dimensi	23	32	14	69
harga	206	212	203	621
kamera	253	86	64	403
layar	88	46	7	141
memori	3	18	1	22
model	10	14	7	31
nfc	20	27	9	56
performa	27	32	16	75
radio	0	2	0	2
ram	61	78	24	163
resolusi	4	10	1	15
sensor	21	21	6	48
sidik jari	0	2	1	3
sim	15	25	3	43
software	25	30	10	65
speaker	12	26	3	41
ukuran	1	2	0	3
usb	2	0	0	2
warna	19	13	1	33
wifi	4	2	0	6

Setelah jumlah sentimen ditemukan, ditetapkan prioritas kebutuhan pelanggan dengan membuat urutan data dengan frekuensi kata kunci paling banyak hingga paling sedikit. Pada Gambar 6, dapat dilihat bahwa kata kunci yang paling banyak memiliki frekuensi adalah “harga” dan yang paling sedikit memiliki frekuensi adalah “usb”.



Gambar 6 Urutan Frekuensi Kata Kunci

Selanjutnya, dilakukan urutan jumlah sentimen negatif mulai dari nilai yang terbesar hingga yang terkecil. Pada Gambar 7, fitur *smartphone* yang paling banyak mendapatkan komentar negatif adalah kamera.



Gambar 7 Urutan Jumlah Komentar Negatif

Pembahasan

Penelitian ini berhasil melakukan pelabelan sentimen dengan bantuan pembelajaran mesin SVM. Proses ini melibatkan beberapa tahap, termasuk preproses data, pembuatan model pembelajaran mesin, dan pelabelan data baru. Setelah data diklasifikasi oleh SVM, dilakukan pemanfaatan hasil analisis sentimen dengan menyaring data berdasarkan kata kunci fitur produk *smartphone*. Hasil penghitungan sentimen dibuat dalam bentuk tabel, yang kemudian digabungkan untuk menyatukan jumlah sentimen pada fitur produk yang sama. Selanjutnya, data disatukan dan direpresentasikan untuk menentukan prioritas fitur dengan visualisasi data hasil sentimen analisis masing-masing fitur produk. Dari visualisasi tersebut, dapat dilihat fitur mana yang paling banyak dibahas dan fitur mana yang memiliki sentimen negatif yang lebih banyak dari sentimen

positif. Informasi ini sangat berguna bagi perusahaan untuk meningkatkan kualitas produk yang akan dirilis untuk generasi produk berikutnya. Misalnya, perusahaan dapat mempertimbangkan untuk melakukan peningkatan pada kualitas kamera dan menawarkan harga yang lebih kompetitif. Selain itu, fitur layar, ram dan chipset juga mendapat komentar negatif yang cukup tinggi, sehingga perlu diperhatikan pula untuk perilsan produk selanjutnya. Secara keseluruhan, penelitian ini menunjukkan bagaimana analisis sentimen dapat digunakan untuk mengidentifikasi kebutuhan pelanggan dan membantu dalam pengembangan produk baru.

Kesimpulan

Mendapatkan informasi mengenai bagian perbaikan pada produk untuk tahap Idea Generation dalam proses *new product development* (NPD) dapat dilakukan melalui analisis sentimen terhadap komentar YouTube. Namun, sebelum dapat memberikan label sentimen data komentar, preproses data yang lebih lanjut diperlukan, termasuk vektorisasi data teks menjadi bentuk numerik. Setelah data komentar menjalani preproses dan vektorisasi, maka pelabelan sentimen dapat dilakukan menggunakan algoritma SVM. Hasil analisis sentimen juga memerlukan penyaringan data yang sesuai dengan kamus kata kunci fitur produk yang relevan, agar proses analisis yang dilakukan merupakan representasi fitur produk itu sendiri. Penyaringan data dilakukan, dan kemudian menghitung jumlah sentimen negatif, positif, dan netralnya. Setelah jumlah sentimen ditemukan, maka dilakukan visualisasi data untuk menemukan fitur produk mana yang paling banyak dibahas, jumlah sentimen negatif yang lebih banyak dari sentimen positif, dan urutan terbesar dari fitur produk yang memiliki jumlah sentimen negatif. Dalam hal ini dapat dikemukakan bahwa untuk perilsan produk selanjutnya, yaitu *smartphone* Redmi Note 12 Pro 5g perlu memperhatikan fitur kamera, harga, layar, ram, dan *chipset* berdasarkan urutan jumlah sentimen negatif 5 terbesar.

BIBLIOGRAFI

- Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics (Switzerland)*, *9*(3). <https://doi.org/10.3390/electronics9030483>
- Frick, B., & Kaimann, D. (2017). The impact of customer reviews and advertisement efforts on the performance of experience goods in electronic markets. *Applied Economics Letters*, *24*(17), 1237–1240. <https://doi.org/10.1080/13504851.2016.1270399>
- Giannakis, M., Dubey, R., Yan, S., Spanaki, K., & Papadopoulos, T. (2022). Social media and sensemaking patterns in new product development: demystifying the customer sentiment. *Annals of Operations Research*, *308*(1–2), 145–175. <https://doi.org/10.1007/s10479-020-03775-6>
- Hamid, F., & ABBASI, M. U. (2020). Competitive Advantage Through New Product Development Capabilities. *Archives of Business Research*, *8*(3), 202–209. <https://doi.org/10.14738/abr.83.7933>
- Huang, T. Y., Chen, W. K., Chen, C. W., & Silalahi, A. D. K. (2022). Understanding How Product Reviews on YouTube Affect Consumers' Purchase Behaviors in

- Indonesia: An Exploration Using the Stimulus-Organism-Response Paradigm. Dalam *Human Behavior and Emerging Technologies* (Vol. 2022). Wiley-Hindawi. <https://doi.org/10.1155/2022/4976980>
- Ireland, R., & Liu, A. (2018). Application of data analytics for product design: Sentiment analysis of online product reviews. *CIRP Journal of Manufacturing Science and Technology*, 23, 128–144. <https://doi.org/10.1016/j.cirpj.2018.06.003>
- Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(6), 1120–1126. <https://doi.org/10.29207/resti.v5i6.3588>
- Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2019). Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining. *Sustainability (Switzerland)*, 11(15). <https://doi.org/10.3390/su11154235>
- kodhai, E., nivetha, B., sriakila, K., & suvalakshmi, G. (2020). Sentiment Analysis for Customer Service. *International Journal of Engineering and Advanced Technology*, 9(4), 585–589. <https://doi.org/10.35940/ijeat.D7287.049420>
- Latuny, W., Lawalata, V. O., Pailin, D. B., & Ohoirenan, R. (2021). Sentiment Analysis of Consumers for Determining the Packaging Features of Eucalyptus Oil Products. *Jurnal Ilmiah Teknik Industri*, 20(1), 71–80. <https://doi.org/10.23917/jiti.v20i1.13461>
- Pradhan, R. (2021). Extracting Sentiments from YouTube Comments. *2021 Sixth International Conference on Image Information Processing (ICIIP)*, 1–4. <https://doi.org/10.1109/ICIIP53038.2021.9702561>
- Rathore, A. K., & Ilavarasan, P. V. (2020). Pre- and post-launch emotions in new product development: Insights from twitter analytics of three products. *International Journal of Information Management*, 50, 111–127. <https://doi.org/10.1016/j.ijinfomgt.2019.05.015>
- Sutrilastyo, S., & Diar Astanti, R. (2021). Lexicon-based Sentiment Analysis for Product Design and Development. *International Journal of Industrial Engineering and Engineering Management*, 3(1), 27–31. <https://doi.org/10.24002/ijieem.v3i1.4351>
- Thirumagal Vijaya, M. (2014). New Product Development. Dalam *International Journal of Science and Research*. www.ijsr.net
- Wilie, B., Vincentio, K., Indra Winata, G., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., Purwarianti, A., & Bandung, I. T. (t.t.). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding*. <https://github.com/annisanurulazhar/absa-playground>

Copyright holder:

Ricky Paian Limbong, Ronsen Purba, Muhammad Fermi Pasha (2024)

First publication right:

Syntax Literate: Jurnal Ilmiah Indonesia

This article is licensed under:

