

## ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA BUKALAPAK MENGUNAKAN METODE ALGORITMA TF-IDF

Pradita Eko Prasetyo Utomo\*, Manaar, Ulfa Khaira, Tri Suratno

Fakultas Sains dan Teknologi, Universitas Jambi  
email: [pradita.eko@unja.ac.id](mailto:pradita.eko@unja.ac.id)

### Abstract

*Bukalapak is one of the Customer-To Customer (C2C) e-commerce models. This model is the most widely applied and found on e-commerce sites in Indonesia. The Customer-To Customer (C2C) market is currently still dominant in Indonesia's online retail market. Data collected from Euromonitor estimates that the C2C market contributed 3% of the retail market in Indonesia in 2017, while the B2C market contributed 1.7%. One text mining analysis is that sentiment analysis can be applied to companies that issue a product or service and provide services to receive opinions (feedback) from consumers for the product. Sentiment analysis is applied to classify positive, negative, and neutral feedback from consumers so as to speed up and simplify the company's task to review their product deficiencies. The researcher conducted further analysis on Bukalapak user reviews to find out how user comments or opinions were on Bukalapak using the TF-IDF Algorithm method. And it can be concluded that based on customer review reviews in Bukalapak have a good rating or perception of this Vans shoe product. Can be seen from the results of Sentiments, Sentiment Visualization and WordCloud Visualization which shows that positive reviews have a higher frequency of 70%.*

**Keywords:** *Sentiment Analysis, TF-IDF Algorithm, preprocessing, e-commerce, customers*

### 1. PENDAHULUAN

Indonesia telah menjadi pasar terbesar e-commerce di Asia Tenggara. Pada 2014, Euromonitor mencatat, penjualan online Indonesia mencapai US\$ 1,1 miliar, lebih tinggi dari Thailand dan Singapura. Namun, jika dibandingkan dengan total perdagangan retail, penjualan e-commerce di Indonesia hanya menyumbangkan 0,07 persen. Artinya, pasar e-commerce Indonesia berpeluang untuk tumbuh semakin besar. Apalagi dengan jumlah penduduk dan tingkat produk domestik bruto (PDB) terbesar di ASEAN. Euromonitor memperkirakan rata-rata pertumbuhan tahunan (CAGR) penjualan online Indonesia selama 2014- 2017 sebesar 38% (Alexa, 2018).

Salah satu situs e-commerce terpopuler yang paling banyak dikunjungi di Indonesia menurut data per Oktober 2017-Januari 2018 adalah Bukalapak. Berdasarkan kunjungan ke situs Bukalapak berasal dari mesin pencari sebesar 34,30% dan total

yang terhubung dengan situs Bukalapak sebesar 31.110.000 (DBS, 2017).

Bukalapak merupakan salah satu model e-commerce Customer-To Customer (C2C). Model ini yang paling banyak diterapkan dan di temukan pada situs e-commerce di Indonesia. Pasar Customer-To Customer (C2C) saat ini masih dominan di pasar ritel online Indonesia. Data yang dihimpun dari Euromonitor memperkirakan pasar C2C menyumbangkan 3% dari pasar ritel di Indonesia pada tahun 2017, sedangkan pasar B2C menyumbangkan 1,7% (Euromonitor, 2017).

Selain itu juga masyarakat Indonesia lebih banyak memilih berbelanja online pada model C2C karena banyak terdapat pilihan produk. Oleh sebab itulah pada penelitian ini akan menganalisis model C2C. Ulasan dari pengguna sering digunakan sebagai alat yang efektif dan efisien dalam menemukan informasi terhadap suatu produk atau jasa. Menurut Sung, bahwa penelitian baru-baru inimenemukan hampir

50% dari pengguna internet bergantung pada rekomendasi word-of-mouth (opini) sebelum menggunakan suatu produk, karena review dari pengguna lain dapat menyediakan informasi terbaru dari produk tersebut berdasarkan perspektif pengguna-pengguna lain yang sudah menggunakan produk tersebut (Fanani, F. 2017).

Tanggapan negatif maupun positif dari pengguna bisa jadi dipengaruhi oleh beberapa hal yang belum menjadi perhatian dari Bukalapak. Hal ini mungkin terjadi karena adanya beberapa faktor yang harus diperbaiki dan belum diketahui oleh Bukalapak. Dengan menggunakan text mining dapat dilihat pembicaraan apa saja yang sering dibahas oleh pengguna.

Salah satu analisis *text mining* yaitu analisis sentimen dapat diaplikasikan pada perusahaan yang mengeluarkan suatu produk dan jasa dan menyediakan layanan untuk menerima pendapat (*feedback*) dari konsumen untuk produk tersebut. Analisis sentimen untuk mengelompokan *feedback* positif, negatif, dan netral dari konsumen sehingga mempercepat dan mempermudah tugas perusahaan untuk meninjau kembali kekurangan produk mereka. Apabila ditemukan adanya sentiment negatif, maka perusahaan dapat dengan cepat mengambil tindakan untuk menanggulangnya.

Oleh sebab itu, dalam penelitian ini dilakukan analisis lebih lanjut mengenai ulasan pengguna Bukalapak untuk mengetahui bagaimana komentar atau opini pengguna terhadap Bukalapak. Berdasarkan alasan tersebut, maka dipilih menggunakan metode *Algoritma TF-IDF*.

## 2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data ulasan atau komentar dari produk bukalapak, terkhususnya produk sepatu vans. Data di ambil menggunakan aplikasi Anaconda menggunakan library jupyter anaconda. Adapun untuk melakukan penelitian ini, metode analisis yang dilakukan adalah sebagai berikut:

### 1) Pengumpulan Data

Pada tahap ini, Pengumpulan data di lakukan dengan mengakses situs penjualan

produk sepatu vans pada bukalapak dengan bantuan akses token agar lebih mudah mendapatkan review terhadap produk. Jumlah data yang di gunakan yaitu 300. Sebagai contoh, mengambil 10 data review pada Gambar 1.

	Nama	Ungkapan	Waktu
0	Teddy Mardina	Gosok semua pokoknya deh... Harga kualitas produ...	2019-09-21T11:26:22.334Z
1	Selwani	makasih ya bos pokoknya sih sampai dgn cepet...	2019-09-30T08:14:26.044Z
2	Rahanyah	Original. Kualitasnya TOP banget, koinnya klati!	2019-09-30T07:52:40.307Z
3	Edie	Terbantu: utk keadilan kali, Udah sering belan...	2019-09-30T07:51:57.088Z
4	Tulalia	Oke banget loh ini, ga nyampe minggu sejam pes...	2019-09-30T07:50:04.692Z
5	Rhsan	Saya senang untuk keadilan kalinya bertelanj...	2019-09-30T07:49:51.536Z
6	Opang	Entah apa yang merasuki otak gue hahaha!	2019-09-28T09:38:36.466Z
7	Hazarah Putri	Barang datang sesuai order dan packagingnya rapi h...	2019-09-25T08:39:36.41Z
8	Puti Ratna	Pengiriman mantap rapi dan cepet. Seller respons...	2019-09-25T08:37:45.594Z
9	Hegar Ahmad	Kualitas pengiriman sangat baik, respons perku...	2019-09-25T08:50:31.388Z

Gambar 1. Data Review Pelanggan

Dari data review tersebut akan dijadikan sample penelitian.

### 2) *Preprocesssing Dataset* (Pembersihan data)

Pada tahap ini dilakukannya proses pembersihan data dimana dalam proses ini data yang diperoleh dalam bentuk teks dilakukan dengan menghilangkan *punctuation, stopwords, spelling correction, tokenization, stemming*, dilakukan perbaikan data dengan mengeliminasi data *Missing Value*. Kemudian melakukan suatu proses data sehingga mendapatkan atribut dan distribusi kelas yang relevan.

### 3) Analisis sentiment

Pada tahap ini, analisis sentimen dilakukan dengan Algoritma TF-IDF.

## 3. HASIL DAN PEMBAHASAN

### a. Klasifikasi Data

Tahapan awal klarifikasi data yaitu preprocessing data. Berikut merupakan langkah yang di lakukan peneliti dalam proses preprocessing data:

#### 1. Hilangkan *Punctuation*

Fungsi menghilangkan *punctuation* yaitu agar tidak mengganggu proses perhitungan dalam penerapan algoritma text mining. contohnya tanda baca dan karakter non-alfabet. Pada Gambar 2 ditampilkan hasilnya.

	Thesaurus	JML_Kata	JML_Morf	Avg_Kata	JML_Thesaurus	JML_Petaling	JML_NomorB	JML_Uppercase
0	good semua polidocya ditfarga kullnya pndak g.	0	0	0.00000	0	0	0	0
1	makasih bos paket nya sih dan cepat sip	10	31	4.20000	2	0	0	0
2	sangat bagus ya buat banget kemasin kider	0	42	0.00000	0	0	0	0
3	berkesan di kenalan full order sangat baik	10	74	0.00000	0	0	0	0
4	ada banget nih ya ga kemasin sangat rapi pndak	11	36	0.00000	2	0	0	0
5	semua kemasin full order sangat baik	14	38	0.07429	2	0	0	0
6	ada apa yang kemasih full order sangat baik	0	40	0.00000	0	0	0	0
7	barang akan masuk order di pndaknya sangat si	12	37	0.00000	0	0	0	0
8	pengiriman sangat rapi sangat order sangat	11	46	0.00000	0	0	0	0
9	kepakian pengiriman sangat baik sangat pndak	10	142	0.00000	0	0	0	0

Gambar 2. Hasil punctuation

2. Hilangkan Stopword.

Fungsi menghilangkan stopwords yaitu untuk menghilangkan kata kata yang tidak penting dalam data. Hasilnya ditunjukkan pada Gambar 3.

	Thesaurus	JML_Kata	JML_Morf	Avg_Kata	JML_Thesaurus	JML_Petaling	JML_NomorB	JML_Uppercase
0	good semua polidocya ditfarga kullnya pndak g.	0	0	0.00000	0	0	0	0
1	makasih bos paket nya sih dan cepat sip	0	41	0.00000	0	0	0	0
2	sangat bagus ya buat banget kemasin kider	0	42	0.00000	0	0	0	0
3	berkesan di kenalan full order sangat baik	10	74	0.00000	0	0	0	0
4	ada banget nih ya ga kemasin sangat rapi pndak	10	36	0.07143	0	0	0	0
5	semua kemasin full order sangat baik	11	34	0.00000	0	0	0	0
6	ada apa kemasih full order sangat baik	7	41	0.00000	0	0	0	0
7	barang akan masuk order di pndaknya sangat si	12	75	0.00000	0	0	0	0
8	pengiriman sangat rapi sangat order sangat	0	36	0.00000	0	0	0	0
9	kepakian pengiriman sangat baik sangat pndak	10	142	0.00000	0	0	0	0

Gambar 3. Hasil menghilangkan stopwords

3. Spelling Correction

Fungsinya yaitu pengecekan ejaan dalam kata baik bahasa inggris maupun bahasa indonesia. Karena banyak kebiasaan user dalam menuliskan sebuah kata dengan menyingkatnya atau typo, jika penulisan singkatan atau typo tidak di perbaiki, ada kemungkinan besar dapat mempengaruhi hasil analisis akhirnya. Maka solusi mengatasi masalah ini dengan melakukan *spelling correction*. Hasilnya ditunjukkan pada Gambar 4.

	Thesaurus	JML_Kata	JML_Morf	Avg_Kata	JML_Thesaurus	JML_Petaling	JML_NomorB	JML_Uppercase
0	good semua polidocya ditfarga kullnya pndak g.	0	0	0.00000	0	0	0	0
1	makasih bos paket nya sih dan cepat sip	0	41	0.00000	0	0	0	0
2	sangat bagus ya buat banget kemasin kider	0	42	0.00000	0	0	0	0
3	berkesan di kenalan full order sangat baik	10	74	0.00000	0	0	0	0
4	ada banget nih ya ga kemasin sangat rapi pndak	10	36	0.07143	0	0	0	0
5	semua kemasin full order sangat baik	11	34	0.00000	0	0	0	0
6	ada apa kemasih full order sangat baik	7	41	0.00000	0	0	0	0
7	barang akan masuk order di pndaknya sangat si	12	75	0.00000	0	0	0	0
8	pengiriman sangat rapi sangat order sangat	0	36	0.00000	0	0	0	0
9	kepakian pengiriman sangat baik sangat pndak	10	142	0.00000	0	0	0	0

Gambar 4. Hasil Spelling Corection

4. Tokenization

*Tokenization* (tokenisasi) yaitu memecahkan kalimat menjadi bagian-bagian kata. Hasil tokenisasi ditunjukkan pada Gambar 5.

Before: makasih bos paket nya sih dan cepat sip  
After: ['makasih', 'bos', 'paket', 'nya', 'sih', 'dan', 'cepat', 'sip']

Gambar 5. Hasil Tokenization

5. Stemming

*Stemming* bertujuan untuk mentransformasikan kata menjadi kata dasarnya atau kata baku. Yang menggunakan library Sastrawi untuk proses stemming bahasa Indonesia. Hasil dari proses stemming diperlihatkan pada Gambar 6.

	Thesaurus	JML_Kata	JML_Morf	Avg_Kata	JML_Thesaurus	JML_Petaling	JML_NomorB	JML_Uppercase
0	good semua polidocya ditfarga kullnya pndak g.	0	0	0.00000	0	0	0	0
1	makasih bos paket nya sih dan cepat sip	0	39	0.00000	0	0	0	0
2	sangat bagus ya buat banget kemasin kider	0	42	0.00000	0	0	0	0
3	berkesan di kenalan full order sangat baik	10	74	0.00000	0	0	0	0
4	ada banget nih ya ga kemasin sangat rapi pndak	10	36	0.07143	0	0	0	0
5	semua kemasin full order sangat baik	11	34	0.00000	0	0	0	0
6	ada apa kemasih full order sangat baik	7	39	0.00000	0	0	0	0
7	barang akan masuk order di pndaknya sangat si	12	75	0.00000	0	0	0	0
8	pengiriman sangat rapi sangat order sangat	0	34	0.00000	0	0	0	0
9	kepakian pengiriman sangat baik sangat pndak	10	142	0.00000	0	0	0	0

Gambar 6. Hasil Stemming

b. Algoritma TF-IDF

1. Term Frequency (TF)

*Term Frequency* adalah frekuensi kemunculan sebuah kata. Semakin besar nilai TF nya, maka semakin besar pula bobot nya. Hasil TF ditunjukkan pada Gambar 7.

Kata	Term Frequency	
14	cepat	7.0
65	di	6.0
58	barang	6.0
7	produk	4.0
67	sewal	4.0
74	sangat	4.0
47	ini	4.0
0	belanja	4.0
8	makasih	3.0
95	ukuran	3.0

Gambar 7. Hasil Term Frequency

2. Inverse Document Frequency (IDF)

IDF menunjukkan hubungan ketersediaan sebuah term dalam seluruh dokumen. Semakin sedikit TF nya, maka nilai IDF semakin besar. Hasil IDF diperlihatkan pada Gambar 8.

Kata	Term Frequency	IDF
14	cepat	7.0 1.734801
65	di	6.0 2.833213
58	barang	6.0 2.833213
7	produk	4.0 2.233775
67	sewal	4.0 2.833213
74	sangat	4.0 2.833213
47	ini	4.0 2.833213
0	belanja	4.0 0.753772
8	makasih	3.0 1.734801
95	ukuran	3.0 2.833213

Gambar 8. Hasil IDF

3. *Term Frequency – Inverse Document Frequency (TF-IDF)*

Algoritma TF-IDF (*Term Frequency – inverse Document Frequency*) adalah salah satu algoritma yang digunakan untuk menganalisa hubungan antara sebuah kalimat dengan sekumpulan dokumen dan metode untuk menghitung bobot setiap kata yang paling umum di gunakan. Rumus umum untuk TF-IDF adalah penggabungan dari formula perhitungan raw TF dan formula IDF dengan cara mengalikan nilai *Term Frequency (TF)* dengan nilai *Inverse Document Frequency (IDF)* yang ditunjukkan pada persamaan (1). Serta hasil dari TF-IDF diperlihatkan pada Gambar 9.

$$w_{ij} = tf_{ij} \times idf_j \quad (1)$$

	Kata	Term Frequency	IDF	TF-IDF
14	sepai	7.0	1.734601	12.142207
65	di	6.0	2.833213	16.999280
58	besar	6.0	2.833213	16.999280
7	produk	4.0	1.223775	4.895102
67	sesuai	4.0	2.833213	11.332853
74	sangat	4.0	2.833213	11.332853
47	ini	4.0	2.833213	11.332853
8	belanja	4.0	0.753772	3.015087
8	makasih	3.0	1.734601	5.203803
95	ukuran	3.0	2.833213	8.499640

Gambar 9. Hasil TF-IDF

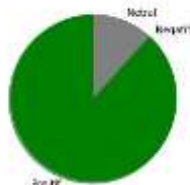
4. Hasil Sentimen, Visualisasi Sentimen dan Visualisasi WordCloud

Hasil Analisis Sentimen ditunjukkan pada Gambar 10 dan Gambar 11 serta telah diberikan pelabelan ke dalam kelas sentimen positif, negatif, dan netral.

Uraian	Jml. Sesi	Jml. Positif	Jml. Negatif	Jml. Netral	Jml. Tidak	Jml. Tidak	Jml. Tidak	Jml. Tidak	Jml. Tidak
1. gant semua pembelian sehingga masalah selesai	3	0	0	3	0	0	0	0	0
2. mantap banget buat aku sih baru coba aja	30	1	0	0	0	0	0	0	0
3. anggap aja ya lah barangnya	4	0	0	0	0	0	0	0	0
4. lumayan ga terlalu bad udah cukup bagus	19	7	0	0	0	0	0	0	0
5. aku harap dia ga nyerah semoga bisa bertahan	14	0	0	0	0	0	0	0	0
6. memang lumayan bagus barangnya sih	11	0	0	0	0	0	0	0	0
7. udah aja masalahnya udah selesai	7	0	0	0	0	0	0	0	0
8. barangnya sesuai order di katalognya sih	12	7	0	0	0	0	0	0	0
9. pengiriman lumayan cepet sih udah sampai	8	0	0	0	0	0	0	0	0
10. lumayan gampang banget masalahnya	10	0	0	0	0	0	0	0	0

Gambar 10. Hasil Sentimen

Hasil analisis sentimen dari 17 ungkapan pengguna



Gambar 11. Hasil Visualisasi Sentimen

Dari Gambar 11 dapat diketahui bahwa hasil pelabelan kelas sentimen menunjukkan bahwa jumlah ulasan positif memiliki frekuensi yang lebih tinggi dibandingkan dengan jumlah ulasan Netral, dan tidak memiliki ulasan Negatif dari 17 ungkapan pengguna.



Gambar 12. Hasil Visualisasi WordCloud

Sementara itu dengan menggunakan Visualisasi Wordcloud yang ditunjukkan pada Gambar 12 memberikan gambaran yang lebih jelas tentang topik dan kata-kata positif yang sering digunakan pengguna Bukalapak dalam memberikan ulasan. Semakin besar ukuran kata pada wordcloud menggambarkan semakin tinggi pula frekuensi kata tersebut, artinya semakin sering pengguna menggunakan kata tersebut sebagai topik pembicaraan atau penilaian positif dalam ulasan. Beberapa topik yang sering di bahas pengguna Bukalapak diantaranya adalah cepat, barang, produk, sesuai, makasih, belanja, sangat dan sebagainya.

5. KESIMPULAN

Berdasarkan hasil analisis menggunakan algoritma TF-IDF dapat di simpulkan bahwa berdasarkan ulasan review pelanggan di bukalapak mempunyai penilaian ataupun persepsi yang baik terhadap produk sepatu ini. Dapat dilihat dari hasil Sentimen, Visualisasi Sentimen dan Visualisasi WordCloud yang menunjukkan bahwa ulasan positifnya memiliki frekuensi yang lebih tinggi yaitu 70%.

6. REFERENSI

Alexa. 2018. *Traffic Statistic Bukalapak.com*. <https://www.alexacom/siteinfo/bukalapak.com>.

- DBS. 2017. *Indonesia Pusat E- commerce ASEAN*. <https://www.dbsinsights.com/id/digital-dan-inovasi/indonesia-pusat-e-commerce-asean>.
- Euromonitor. 2017. *Market Research Indonesia*.<http://www.euromonitor.com/in-donesia>.
- Fanani, F. 2017. *Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen*. Skripsi. Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta.