

ANALISIS SENTIMEN UNTUK MENGETAHUI TINGKAT POPULARITAS SITUS BELANJA ONLINE DI INDONESIA MENGGUNAKAN METODE NAIVE BAYES

Okky Zaenur Endrawan✉, Walid

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia
Gedung D7 Lt. 1, Kampus Sekaran Gunungpati, Semarang, 50229

Info Artikel

Sejarah Artikel:
Diterima
Disetujui
Dipublikasikan

Keywords:

Tokopedia, Shopee, Bukalapak, naive bayes, sentiment analysis

Abstrak

Tokopedia, Shopee, dan Bukalapak merupakan perusahaan yang terlibat dalam pasar dan toko *online*. Untuk menanggapi pendapat, kritik dan keluhan, Tokopedia, Shopee, Bukalapak memiliki suatu *platform website* yang merupakan tempat untuk mengakses dengan aplikasi mobile dan juga bisa untuk mengulas layanan dan jasa yang ditawarkan oleh ketiga perusahaan tersebut. *Platform website* tersebut bernama *Google Play Store*. Jenis ulasannya merupakan sebuah rating dalam berbentuk bintang dengan skala 1-5 dan teks yang tidak terstruktur dalam jumlah besar. Kondisi ini dapat menyebabkan perusahaan belanja *online* melewatkan informasi berguna dari koleksi dokumen teks. Analisis setimen ditetapkan dalam studi kasus ini untuk melakukan pemeringkatan popularitas situs belanja *online* di Indonesia. Pada penelitian ini hanya membahas tiga *online Shop*, yakni Tokopedia, Shopee, Lazada. Studi kasus ini menggunakan metode *Naïve Bayes Classifier* (NBC) dengan pengujian *10-fold cross validation* untuk memastikan keakuratan data. Metode NBC digunakan karena lebih cocok dalam pengoperasian analisis sentiment pada penelitian ini dibandingkan dengan metode lainnya. Pada penelitian ini Tokopedia merupakan *online shop* yang paling populer dibandingkan dengan Shopee dan Bukalapak. Berdasarkan perhitungan pemeringkatan popularitasnya didapatkan untuk *online shop* Tokopedia memiliki tingkat popularitas 75%, Shopee 52% dan Bukalapak 87%. Dengan uji validitas metode NBC menggunakan *10-fold cross validation* didapatkan akurasi untuk Tokopedia 92,5%, Shopee 67,5% dan Bukalapak 95,8%.

Abstract

Tokopedia, Shopee, and Bukalapak are the companies involved in the online market and shop. To respond to opinions, criticisms and complaints, Tokopedia, Shopee, Bukalapak have a website platform which is a place to access with a mobile application and also to review the services and services offered by the three companies. The website platform is called the Google Play Store. The comments type is a star rating on a scale of 1-5 and large amounts of unstructured text. This condition can cause online shopping companies to miss useful information from collections of text documents. Setiment analysis is defined in this case study to rank the popularity of online shopping sites in Indonesia. In this study only discussed three online shops, namely Tokopedia, Shopee, Lazada. This case study uses the Naïve Bayes Classifier (NBC) method with 10-fold cross validation testing to ensure data accuracy. The NBC method was used because it is more suitable for the operation of sentiment analysis in this study compared to other methods. In this study, Tokopedia is the most popular online shop compared to Shopee and Bukalapak. Based on the calculation of the popularity ranking, it was found that the online shop Tokopedia had a popularity rate of 75%, Shopee 52% and Bukalapak 87%. By testing the validity of the NBC method using 10-fold cross validation, the accuracy for Tokopedia 92.5%, Shopee 67.5% and Bukalapak 95.8%.;

PENDAHULUAN

Menurut Lembaga web statistic internet *statista.com*, populasi *netizen* tanah air mencapai 184,94 juta orang pada 2019. Angka yang berlaku untuk setiap orang yang mengakses internet setidaknya satu kali setiap bulan itu mendudukkan Indonesia di peringkat ke-4 terbesar di dunia dalam hal jumlah pengguna internet. Pada 2020, *statista* memperkirakan *netizen* Indonesia bakal mencapai 199,16 juta orang, mengalahkan. Jepang di peringkat yang pertumbuhan jumlah pengguna internetnya lebih lambat. Secara keseluruhan, jumlah pengguna internet di seluruh dunia diproyeksikan bakal mencapai 4 miliar orang pada 2018. Tiga tahun setelahnya, pada 2018, diperkirakan sebanyak 4,9 miliar manusia di bumi bakal mengakses internet setidaknya sekali tiap satu bulan. (Statista, 2019)

Definisi *marketplace* berdasarkan ilmu *One Data* adalah fasilitator pembelanjaan *online* yang tidak memiliki inventarisasi barangnya sendiri. Sedangkan, *e-commerce* merupakan istilah yang digunakan untuk menyebut suatu kegiatan perdagangan melalui sistem elektronik seperti internet ataupun televisi. Perkembangan teknologi yang menyebabkan pertumbuhan pengguna internet *mobile* semakin tinggi berdampak juga ke jumlah aplikasi yang tersedia untuk perangkat *smartphone*. Terdapat dua penyedia aplikasi untuk dua jenis *smartphone* yang berbeda yaitu *Google Play Store* dan *Apple Store*. (Inews.id, 2020) Pengguna *android* dapat mengakses dan mengunduh berbagai aplikasi melalui *Apple Store* sedangkan pengguna *android* melalui *Google Play Store*. Saat ini jumlah aplikasi yang tersedia di *Google Play Store* sebanyak 3,6 juta aplikasi, data tersebut adalah data hingga Maret 2018. Pada awalnya layanan ini diluncurkan pada Oktober 2008 dengan nama *Android Market*. Berbagai jenis aplikasi tersedia disana mulai dari permainan, buku, aplikasi chat, musik, majalah dan TV. Sebagian besar aplikasi yang tersedia di layanan tersebut adalah

gratis, namun para perusahaan pembuat aplikasi memiliki model bisnis tersendiri untuk tetap mendapatkan keuntungan. (Cnnindonesia.com, 2019)

Pada penelitian ini, *sentiment analysis* akan digunakan untuk menentukan peringkat popularitas tujuan pembelian berdasarkan data ulasan atau opini yang terdapat di toko media digital yang juga terdapat situs belanja *online*. Manfaat yang dapat diambil pada penelitian ini, yakni pengguna aplikasi belanja *online* dapat lebih bijak dalam menentukan situs mana yang aman dalam proses transaksinya. Selain itu, pihak pemilik situs juga dapat meningkatkan kinerja layanannya dengan adanya analisis ini. Pada penelitian ini juga, peneliti memutuskan untuk menggunakan strategi klasifikasi berdasarkan pada algoritma *Naive Bayes* (NB) karena merupakan suatu metode yang sederhana dan merupakan dan intuitif yang kinerjanya mirip dengan pendekatan lain. Selain itu, berdasarkan penelitian-penelitian terdahulu, NB menggabungkan efisiensi (waktu kinerja yang optimal) dengan cukup akurat.

METODE

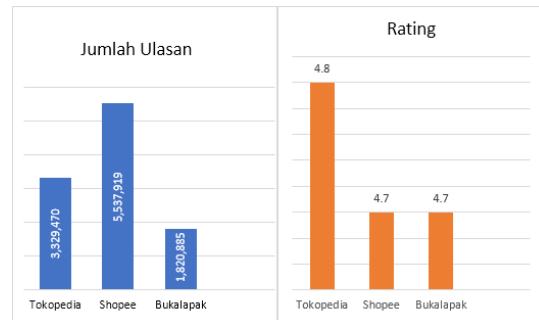
Pengambilan data pada penelitian ini didapatkan dari ulasan para pengguna aplikasi Tokopedia, Shopee, dan Bukalapak di *Google Play Store*, dengan menggunakan teknik *web scraping* (*add-on* dari *Software Chrome* yakni *Web Scraper*). Dataset yang diambil sebanyak masing-masing 120 ulasan dengan total menjadi 360 ulasan, jumlah data seimbang (*balanced*) setiap labelnya, karena dengan data yang tidak seimbang (*imbalanced*), klasifikasi yang dibangun memiliki kecenderungan untuk mengabaikan *minority*. Data yang telah diperoleh kemudian dilakukan *text preprocessing*. Hal pertama yang dilakukan dalam *text preprocessing* adalah *case folding* yaitu proses penyamaan *case* pada dokumen menjadi *lowercase*, kemudian dilakukan *tokenizing* yaitu proses pemecahan

dokumen menjadi kata perkata, lalu dilakukan proses *filtering* yaitu proses penghapusan kata yang tidak diperlukan pada dokumen, dan yang terakhir adalah *stemming* yaitu mencari kata dasar pada tiap kata dalam dokumen. Setelah dilakukan *text preprocessing* dilakukan pelabelan data. Pelabelan data dilakukan dengan menghitung jumlah kata positif dan negatif pada tiap tanggapan. Apabila jumlah kata positif lebih besar daripada jumlah kata negatif maka tanggapan akan dikategorikan menjadi tanggapan positif, begitu juga sebaliknya. Data yang sudah dikategorikan, kemudian dilakukan pengklasifikasian menggunakan *naive bayes*. *Naive Bayes Classifier* merupakan teknik pembelajaran algoritma *data mining* yang memanfaatkan metode probabilitas dan statistik. *Naive Bayes Classifier* dalam melakukan klasifikasi terdapat dua proses penting yaitu *learning (training)* dan *testing*. *Naive Bayes Classifier* bekerja sangat baik dibanding dengan model *classifier* lainnya. Hal ini dibuktikan pada jurnal Xhemali, Daniela, Chris J. Hinde, and Roger G. Stone. "Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages."(Xhemali, J. Hinde, & G. Stone, 2009), mengatakan bahwa "Naive Bayes Classifier memiliki tingkat akurasi yg lebih baik dibanding model *classifier* lainnya". Algoritma tersebut akan dilihat performa kinerjanya dalam mengklasifikasikan data dengan evaluasi model *10-fold cross validation*. Penggunaan *10-fold cross validation* untuk menghilangkan bias pada data. Pelatihan dan pengujian pengujian dilakukan sebanyak 10 kali. Pada percobaan pertama, *subset* S1 diperlakukan sebagai data pengujian dan *subset* lainnya diperlakukan sebagai data pelatihan, pada percobaan kedua *subset* S1, S3, ... Sk menjadi data pelatihan dan S2 menjadi data pengujian, dan seterusnya. Visualisasi data tanggapan dilakukan dengan *wordcloud* dan asosiasi kata. *Wordcloud* merupakan sebuah sistem yang memunculkan visualisasi kata-kata dengan memberikan penekanan pada frekuensi kemunculan kata terkait dalam wacana tertulis.

HASIL DAN PEMBAHASAN

Data ulasan yang telah diperoleh kemudian dilakukan pengelompokan data secara manual, selanjutnya dilakukan pengumpulan data terhadap jumlah *rating* dari konsumen yang ditujukan kepada situs

belanja Tokopedia, Shopee, dan Bukalapak. Gambar dibawah menunjukkan jumlah rating dari konsumen pada bulan Oktokber tahun 2020.



Gambar 1. jumlah ulasan dan rating di *Google Play Store* tahun 2020

Berikut merupakan prosentase ulasan positif maupun negatif dari *online shop* Tokopedia, Shopee, dan bukalapak berdasarkan perhitungan manual. Perhitungan manual dilakukan oleh peneliti untuk melakukan perbandingan dengan hasil perhitungan menggunakan aplikasi *Rapidminer* untuk menghasilkan data yang lebih akurat. Perhitungan ini dilihat berdasarkan ulasan pengguna.

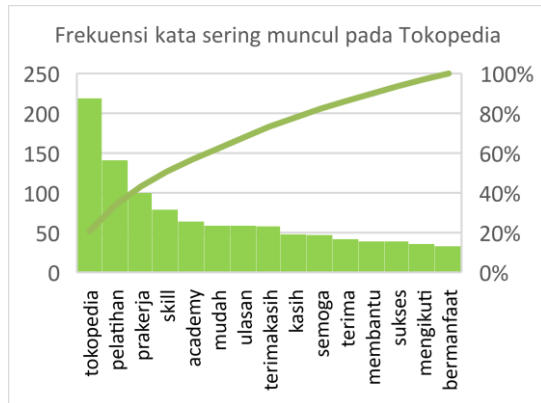
Apabila ulasan yang ditulis mengandung kata-kata positif dan rating yang bagus, maka ulasan tersebut masuk kedalam sentimen positif, begitu pula sebaliknya

Tabel 1. Presentase Sentimen *Online Shop*

No.	Situs Belanja	Jumlah Sentimen		Persentase Sentimen	
		Positif	Negatif	Positif	Negatif
1	Tokopedia	90	30	75%	25%
2	Shopee	62	58	52%	48%
3	Bukalapak	104	16	87%	13%

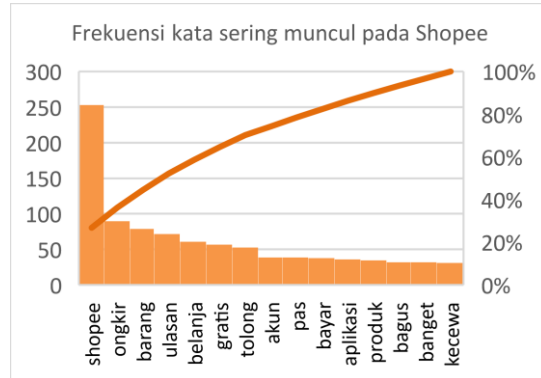
Sebelum dilakukan analisis, langkah selanjutnya yang akan dilakukan adalah pra proses pada data ulasan di *Google Play Store* yang memuat ulasan ketiga situs belanja online, yakni Tokopedia, Shopee, dan Bukalapak terdiri dari beberapa tahapan, yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Proses *case folding* adalah mengubah data teks menjadi *lowercase* dengan tujuan agar kata yang sama namun berbeda secara penulisan huruf kapital dan tidak, tidak dianggap kata yang berbeda. Setelah melakukan proses *case folding* pada data teks, langkah selanjutnya adalah proses *tokenizing* dimana proses memecah kalimat

menjadi kata-kata. Setelah melakukan proses *tokenizing* pada data teks, langkah selanjutnya adalah proses *filtering*. *Filtering* pada data ulasan merupakan proses membersihkan ulasan dari kata yang tidak diperlukan untuk mengurangi noise. Kata yang dihilangkan dalam ulasan adalah karakter HTML, *emoticons*, *hashtag* (#), *username* (@username), URL, dan menghilangkan *stopword*. Proses *stemming* dilakukan untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan, dan *confixes* (kombinasi dari awalan dan akhiran). Pada penelitian ini kamus *stemming* yang digunakan adalah *Liu's word list*. Hasil dari proses *stemming* dipakai sebagai kata kunci dari data ulasan yang akan dianalisis dengan menggunakan metode NBC (*Naive Bayes Classification*). Frekuensi setiap kata dalam ulasan Tokopedia, Shopee, dan Bukalapak ditunjukkan pada Gambar 2, Gambar 3 dan Gambar 4.



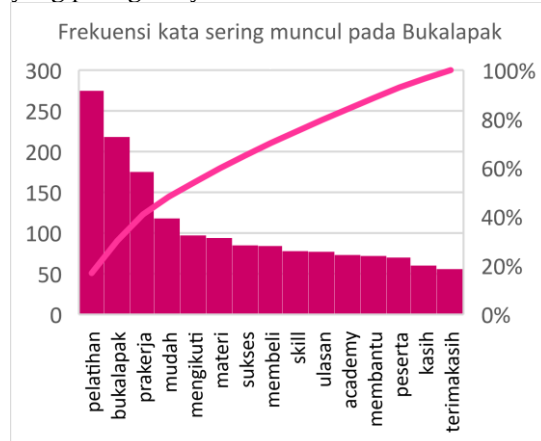
Gambar 2. Frekuensi kata yang sering muncul pada Tokopedia.

Berdasarkan Gambar 2 dapat diketahui frekuensi dari 15 kata tertinggi situs belanja *online* Tokopedia. Kata “pelatihan” merupakan kata yang paling banyak dibicarakan oleh konsumen, sedangkan kata “prakerja” merupakan kata yang paling banyak kedua dibicarakan oleh konsumen. Hal tersebut dikarenakan bahwa Tokopedia kemungkinan pada saat ini banyak menjual jasa pelatihan daripada barang. Kata “skill” merupakan kata berikutnya yang paling banyak dibicarakan.



Gambar 3. Frekuensi kata yang sering muncul pada Shopee

Berdasarkan Gambar 3 dapat diketahui frekuensi dari 15 kata tertinggi situs belanja *online* Shopee. Kata “ongkir” merupakan kata yang paling banyak dibicarakan oleh konsumen, sedangkan kata “barang” merupakan kata yang paling banyak kedua dibicarakan oleh konsumen. Hal tersebut dikarenakan bahwa Shopee adalah *market place* yang banyak menjual barang daripada jasa. Kata “ulasan” merupakan kata berikutnya yang paling banyak dibicarakan.



Gambar 4. Frekuensi kata yang sering muncul pada Bukalapak

Berdasarkan Gambar 4 dapat diketahui frekuensi dari 15 kata tertinggi situs belanja *online* Bukalapak. Kata “pelatihan” merupakan kata yang paling banyak dibicarakan oleh konsumen, sedangkan kata “prakerja” merupakan kata yang paling banyak kedua dibicarakan oleh konsumen. Hal tersebut dikarenakan bahwa Tokopedia kemungkinan pada saat ini banyak menjual jasa pelatihan daripada barang. Kata “mudah” merupakan kata berikutnya yang paling banyak dibicarakan.

Naive Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema

Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat (*naïf*) akan independensi dari masing-masing kondisi/kejadian.

Tabel 2. Hasil Perhitungan NBC Pada *Online Shop*.

No.	Situs Belanja	Jumlah Sentimen		Persentase Sentimen	
		Positif	Negatif	Positif	Negatif
1	Tokopedia	87	33	72%	28%
2	Shopee	53	67	44%	56%
3	Bukalapak	105	15	87%	13%

Pengujian validitas data pada penelitian ini menggunakan *tools RapidMiner* untuk menghasilkan suatu output klasifikasi dan keakurasian data. *10-fold cross validation* digunakan dalam penelitian ini untuk menguji keakuratan data.

Tabel 3. Tingkat akurasi *Naïve Bayes* pada setiap *Online Shop*

Rank	Situs	Tingkat Akurasi
1	Bukalapak	95,83%
2	Tokopedia	92,50%
3	Shopee	67,50%

Tabel diatas menjelaskan bahwa secara tingkat akurasi Bukalapak menempati peringkat 1 untuk *Online Shop* pada tahun 2020. Kemudian disusul dengan Tokopedia dan Shopee.

Pendapat dari konsumen belanja *online* di Tokopedia, Shopee dan Bukalapak dapat divisualisasikan dalam bentuk *Wordcloud*. *Wordcloud* merupakan representasi grafis dari sebuah dokumen yang dilakukan dengan plotting kata-kata yang sering muncul pada sebuah dokumen pada ruang dua dimensi. Frekuensi dari kata yang sering muncul ditunjukkan melalui ukuran huruf kata tersebut. Semakin besar ukuran kata menunjukkan semakin besar frekuensi kata tersebut muncul dalam dokumen. Berikut adalah tampilan visualisasi data dengan

Wordcloud berdasarkan hasil uji yang telah dilakukan



Gambar 5. Visualisasi *Wordcloud* Tokopedia.

Berdasarkan Gambar 5 dapat diketahui bahwa tiga kata yang sering muncul pada akun Tokopedia adalah tokopedia, pelatihan, dan prakerja. Peningkatan pelayanan dan respon positif kepada pengguna terhadap pelatihan prakerja dari *online shop* Tokopedia perlu dipertahankan dan diberi apresiasi. Selain respon positif dari pengguna, respon negatif juga perlu diketahui oleh Tokopedia untuk meningkatkan pelayanan konsumen.



Gambar 6. Visualisasi *Wordcloud* Shopee

Berdasarkan Gambar 6 dapat diketahui bahwa tiga kata yang sering muncul pada akun Shopee adalah shopee, ongkir, dan barang. Peningkatan pelayanan dan respon positif kepada pengguna terhadap ongkos kirim suatu barang dari *online shop* Shopee perlu dipertahankan dan diberi apresiasi. Selain respon positif dari pengguna, respon negatif juga perlu diketahui oleh Shopee untuk meningkatkan kepuasan dalam berbelanja.



Gambar 7. Visualisasi *Wordcloud* Bukalapak

Berdasarkan Gambar 7 dapat diketahui bahwa tiga kata yang sering muncul pada akun Bukalapak adalah pelatihan, bukalapak dan prakerja. Peningkatan pelayanan dan respon positif kepada pengguna terhadap pelatihan prakerja dari *online shop* Bukalapak perlu dipertahankan dan diberi apresiasi. Selain respon positif dari pengguna, respon negatif juga perlu diketahui oleh Bukalapak untuk meningkatkan pelayanan konsumen.

SIMPULAN

Berdasarkan hasil analisis yang telah diperoleh, kesimpulan yang didapatkan dari penelitian ini yaitu konsumen yang memberikan tanggapan, pendapat, kritik, saran dan masalah *complain* lebih banyak ditujukan pada online shop Shopee dari pada Tokopedia dan Bukalapak. Tingkat akurasi data pada *online shop* Shopee kurang begitu baik, sehingga memungkinkan untuk menggunakan uji yang lain seperti *Random Forest* ataupun *Support Vector Machine*.

Online Shop yang paling populer menurut penelitian ini adalah Bukalapak dengan rata-rata penghitungan manual, dengan nilai sentiment positif 87%, disusul dengan Tokopedia dengan nilai sentimen positif 75%, kemudian di posisi paling akhir adalah Shopee dengan nilai sentiment positif 52%.

Dengan uji validitas metode NBC menggunakan *10-fold cross validation* didapatkan tingkat akurasi untuk Tokopedia 92,50%, Shopee 67,50% dan Bukalapak 95,83%. Artinya, metode NBC tidak sepenuhnya memiliki tingkat akurasi tinggi dalam klasifikasi sentimen analisis. Berdasarkan visualisasi wordcloud, penulis mengambil 3 ulasan teratas yang banyak dibicarakan oleh konsumen pada online shop :a) Tokopedia : pelatihan, prakerja, skill; b) Shopee : ongkir, barang, ulasan; c) Bukalapak : pelatihan, prakerja, mudah

Pada penelitian ini Tokopedia merupakan *online shop* yang paling populer dibandingkan dengan Bukalapak dan Shopee. Berdasarkan perhitungan pemeringkatan keseluruhan didapatkan untuk online shop Tokopedia mendapatkan ranking 1, disusul dengan Shopee dan Bukalapak.

DAFTAR PUSTAKA

- Cnnindonesia.com. (2019). Daftar Toko Online yang Paling Menguasai Pasar RI Selama 2019. Retrieved August 25, 2020, from <https://www.cnnindonesia.com/teknologi/20200124205259-206-468540/daftar-toko-online-yang-paling-menguasai-pasar-ri-selama-2019>
- Inews.id. (2020). 10 E-Commerce Terpopuler di Indonesia. Retrieved August 25, 2020, from <https://www.inews.id/finance/bisnis/10-e-commerce-terpopuler-di-indonesia-tokopedia-terdepan-shopee-geser->

bukalapak

- Statista. (2019). Jumlah pengguna e-commerce Indonesia. Retrieved from <https://www.statista.com/outlook/243/120/ecommerce/indonesia#market-users>

- Xhemali, D., J. Hinde, C., & G. Stone, R. (2009). Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *International Journal of Computer Science*, 4(1), 16–23. Retrieved from <http://cogprints.org/6708/>