

SENTIMEN ANALYSIS TERHADAP NILAI KEPERCAYAAN SEBUAH *ONLINE SHOP* DI INSTAGRAM

Yan Puspitarani

Universitas Widyatama
Jalan Cikutra No. 204A Bandung
yan.puspitarani@widyatama.ac.id

Abstrak

Instagram menjadi tempat yang menarik untuk memasarkan produk. Dengan modal foto dan caption yang menarik, *online shop* akan mendapatkan *followers* yang kemungkinan besar tertarik untuk membeli produk mereka. Akan tetapi, seiring dengan mudahnya membuat akun Instagram, tingkat kepercayaan pembeli terhadap *online shop* tersebut menjadi hal yang paling penting. Oleh karena itu, diperlukan review terhadap *online shop* tersebut. Salah satu caranya adalah dengan memanfaatkan sentimen analysis terhadap komentar-komentar pada foto dan caption di akun Instagram *online shop* yang bersangkutan. Model sistem untuk proses review menggunakan *sentiment analysis* pun diusulkan pada paper ini. Proses *sentiment* akan menggabungkan pendekatan *lexicon* dan *machine learning*. Data yang akan digunakan adalah komentar-komentar terhadap foto dan caption beberapa *online shop* di Instagram.

Kata Kunci :

sentiment analysis, lexicon analysis, machine learning for sentiment

Abstract

Instagram became an attractive place to market the product. With attractive photos and captions, online shop will gain followers who are interested in buying their products. However, Instagram account can create easily so the level of buyer confidence to shop online has become the most important thing. Therefore, the review of the online shop is required. The review can be used by sentiment analysis to the comments of photos and captions in the online shop Instagram account. A model system for the review process using sentiment analysis was proposed in this paper. Sentiment process will combine lexicon and

machine learning approaches. Data will be used in this study are comments on photos and captions several online shop on Instagram.

Keywords:

sentiment analysis, lexicon analysis, machine learning for sentiment

I. PENDAHULUAN

Media sosial telah menjadi trend yang paling banyak dinikmati pengguna Internet untuk berbagi informasi kepada masyarakat luas. Mereka dapat berbagi menggunakan berbagai macam content seperti gambar, video, atau artikel di media sosial. Mereka pun mengekspresikan pendapat atau perasaannya melalui media sosial, seperti *twitter*, *facebook*, Instagram dan sebagainya.

Salah satu media sosial yang cukup pesat peningkatan penggunaannya adalah Instagram. Dalam situsnya, Instagram menyatakan jumlah pengguna aktif di situs mereka mencapai 300 juta user per bulannya, dengan rata-rata 70 juta foto diupload setiap hari dan 2.5 milyar likes setiap harinya. Hal ini pun dimanfaatkan *Online Shop* sebagai tempat untuk memasarkan produk mereka. Dengan modal foto dan caption yang menarik, *online shop* akan mendapatkan *followers* yang kemungkinan besar tertarik untuk membeli produk mereka.

Namun demikian, dengan banyaknya *online shop* di Instagram, tingkat kepercayaan pembeli terhadap *online shop* tersebut menjadi hal yang paling penting. Customer sebaiknya mencari informasi *online shop* lebih lanjut sebelum memutuskan untuk membeli. Informasi tersebut didapatkan lewat profil Instagram, *comment*, *testimonial*, dan *followers*.

Komentar dan testimonial di Instagram dapat dimanfaatkan untuk menilai tingkat kepercayaan sebuah *online shop* di Instagram. Nilai kepercayaan tersebut didapatkan dengan melihat komentar positif atau negatif dari setiap gambar dan *caption* yang diposting sebuah *online shop* di Instagram. Penentuan komentar positif atau negatif ini merupakan salah satu bagian *sentiment analysis*.

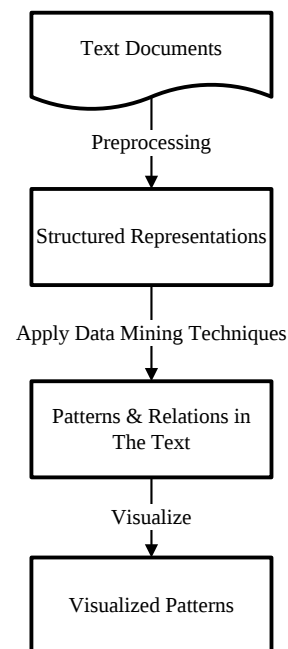
Sentiment Analysis, atau disebut juga opinion mining, merupakan bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap dan emosi publik terhadap suatu entitas seperti produk, pelayanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya (Liu, 2012). *Sentiment analysis* memiliki banyak kegunaan, salah satu kegunaannya adalah menilai opini publik mengenai suatu produk. Dari sisi bisnis, penilaian masyarakat terhadap produk yang mereka jual sangat berguna untuk penjualan mereka.

Terdapat beberapa penelitian terkait hal ini, yaitu Xing Fang melakukan analisis sentimen terhadap review produk dari Amazon menggunakan metode *Naïve Bayesian*, *Random Forest*, and *Support Vector Machine* sebagai metode klasifikasinya (Fang, et.al, 2015). Kemudian, Sonali D Ingale juga melakukan *sentiment analysis* terhadap review mobile, camera, printer dan tv. Mereka mengklasifikasikan review terhadap produk tersebut menjadi positif, negatif dan netral (Ingale, 2015). Selain itu, T.K. Das mengimplementasikan sebuah framework untuk melakukan *sentiment analysis* terhadap opini tentang Samsung Galaxy di twitter (Das, et.al, 2014). Hal yang sama pun dilakukan M.M Mostafa terhadap opini para pengguna produk dengan merk terkenal seperti Nokia, Samsung, IBM, dan Egypt Air. Hasil penelitian menunjukkan bahwa perusahaan dapat secara efektif menggunakan *blogosphere* untuk mendesain ulang kampanye pemasaran dan iklan mereka (Mostafa, 2013). Selain review produk, *sentiment analysis* juga dilakukan Asur untuk memperkirakan pendapatan suatu film dengan menganalisis data dari twitter (Sitaram, et.al, 2010).

Berdasarkan penelitian-penelitian yang sudah dilakukan tersebut, pada penelitian ini akan dilakukan review terhadap suatu *online shop* yang ada di Instagram untuk menentukan tingkat kepercayaan konsumen terhadap *online shop* tersebut. Review dilakukan menggunakan *sentiment analysis* terhadap komentar-komentar yang ada di *online shop* tersebut.

II. TEXT MINING

Text mining adalah salah satu bidang khusus dari *data mining*. Hanya saja, yang membedakannya adalah pada sumber datanya, dimana text mining bersumber dari kumpulan dokumen atau teks. Sesuai dengan buku *The Text Mining Handbook*, text mining dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam *data mining* yang salah satunya adalah klasifikasi. Pada proses klasifikasi ini, dokumen akan dikelompokkan ke dalam kategori yang sesuai berdasarkan pola yang dibuat pada proses *training*.



Gambar 1 Proses Text Mining[13]

Sebelum dilakukan proses klasifikasi, perlu dilakukan tahapan text preprocessing terlebih dahulu untuk mengubah bentuk text yang belum terstruktur menjadi bentuk yang sesuai untuk inputan klasifikasi.

Text Preprocessing

Pada text mining, informasi yang akan digali berisi informasi-informasi yang strukturnya sembarang. Oleh karena itu, diperlukan proses pengubahan bentuk menjadi data yang terstruktur sesuai kebutuhannya untuk proses dalam *data*

mining, yang biasanya akan menjadi nilai-nilai numerik. Proses ini sering disebut Text Preprocessing. Setelah data menjadi data terstruktur dan berupa nilai numerik maka data dapat dijadikan sebagai sumber data yang dapat diolah lebih lanjut.

Text Preprocessing dapat juga dikatakan sebagai feature generation atau pembentukan feature dari dokumen mentah menjadi data yang terstruktur. Feature generation dilakukan dengan melakukan *tokenization* terhadap dokumen yang menghasilkan kumpulan *token* kemudian disaring dengan membuang *token-token* yang ada dalam daftar *stopword* yang selanjutnya menjadi *term-term* atau *feature-feature* yang hanya berupa sekumpulan kata-kata. Setelah itu, dilakukan penyusunan dokumen dimana setiap *term* hasil pemisahan tersebut muncul di dalamnya. Hasil penyusunan dokumen berupa frekuensi kemunculan setiap *term* pada dokumen tersebut (Das, et.al, 2014).

Hasil penyusunan dokumen tersebut direpresentasikan dalam bentuk matriks dengan field berupa id dokumen, frekuensi kemunculan setiap *term* atau bisa dikatakan sebagai *feature*, dan kategori. Matriks ini selanjutnya dapat digunakan sebagai masukan terhadap klasifikasi.

III. SENTIMENT ANALYSIS

Sentiment Analysis, atau disebut juga *opinion mining*, merupakan bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap dan emosi publik terhadap suatu entitas seperti produk, pelayanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya (Liu, 2012). *Sentiment Analysis* tidak berbeda jauh dari *text Mining*, yang membedakannya hanya terletak dari klasifikasi yang dihasilkan. Berikut ini adalah contoh opini yang terdapat dalam akun Instagram sebuah *online shop* :

Komentar-komentar yang ada pada gambar 2 merupakan komentar negatif dan positif terhadap dua buah *online shop* yang berbeda. Berdasarkan komentar tersebut, dapat diketahui nilai kepercayaan terhadap *online shop* tersebut. Melalui proses *sentiment analysis*, dapat diketahui nilai positif dan negatif terhadap sebuah *online shop*.



Gambar 2 Komentar negatif (kiri) dan komentar positif (kanan) di dua online shop yang berbeda

Sentiment Analysis dapat dilakukan menggunakan dua buah pendekatan, yaitu :

1. *Lexicon Based Analysis* (Taboada, et.al., 2011).

Pendekatan ini menggunakan sekumpulan kata-kata opini positif dan negatif untuk mengukur suatu kalimat itu positif, negatif, atau netral. Pendekatan ini sangat populer dan membutuhkan fungsi skoring untuk menghitung nilai setiap kalimat berdasarkan kemunculan kata-kata positif atau negatif pada *lexicon*.

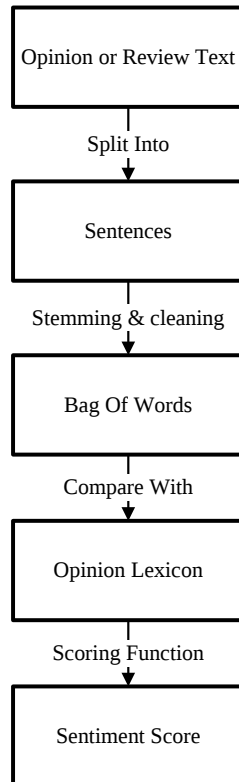
2. *Machine Learning Based Analysis* (Rain, 2013)

Pendekatan ini menggunakan teknik-teknik *machine learning* untuk melakukan training terhadap *classifier* menggunakan opini-opini yang sudah diklasifikasikan sebagai training set. Kemudian menggunakan *classifier* yang sudah di train tersebut untuk menentukan opini yang baru sebagai opini yang positif, negatif atau netral.

III.1 Polarity of Words

Untuk melakukan *sentiment analysis*, diperlukan *opinion lexicon* dan *scoring function* untuk menetapkan nilai terhadap suatu komentar. (Younis, 2015). *Lexicon* ini adalah daftar kata-kata atau daftar makna, dimana setiap inputannya dilabeli dengan nilai polaritasnya (Santos, 2012). Sebagai contoh, walaupun tanpa mengetahui konteksnya, kata-kata seperti cinta, damai, senang, dan sukses

biasa diberi label positive prior polarity, sedangkan kata-kata benci, perang, bosan, dan gagal diberi label negative prior polarity. Sementara itu, kata-kata seperti orang, tabel, dan pohon diberi label netral.



Gambar 3 Pendekatan Lexicon Based Sentiment Analysis (Younis, 2015)

Berikut ini adalah scoring function untuk penentuan nilai sentimen suatu komentar (Younis, 2015):

$$SentimentScore = \sum positiveWords - \sum negativeWords$$

Persamaan (1)

Keterangan :

- Netral : Nilai SentimentScore akan 0, jika kata-kata positif dan negatif jumlahnya sama atau tidak ada kemunculan kata-kata positif dan negatif
- Positive Polarity : Nilai SentimentScore akan positif, jika jumlah kata-kata

positif > jumlah kata-kata negatif

Negative Polarity : Nilai SentimentScore akan negatif, jika jumlah kata-kata negatif > jumlah kata-kata positif

III.2 Document Sentiment Classification

Berdasarkan survey yang dilakukan G. Vinodhini, algoritma klasifikasi pada sentiment analysis bergantung terhadap topiknya. Dalam surveynya terhadap algoritma klasifikasi seperti Naive Bayes, SVM, Hybrid, dan n-gram, tidak ada satu pun yang konsisten menghasilkan performansi bagus. Semua bergantung terhadap feature yang dihasilkan (Vinodhini & Chandrasekaran, 2012)

Oleh karena itu, dalam penelitian ini akan menggunakan Naive Bayes sebagai classifiernya karena Naive Bayes merupakan algoritma yang simple tetapi robust (Rain, 2013).

Naïve Bayes Classifier (NBC) merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan (independent) yang tinggi (Younis, 2015).

Naïve Bayes dikatakan naïve karena penggunaannya yang memakai teorema Bayes dan mengasumsikan ketidakbergantungan dengan naïve (kuat) dimana kemunculan atau ketidakmunculan suatu feature tidak berhubungan dengan kemunculan atau ketidakmunculan feature-feature yang lainnya. Naïve Bayes bekerja sangat baik ketika diuji terhadap dataset yang sebenarnya, terutama ketika dikombinasikan dengan beberapa prosedur pemilihan atribut (Written & Frank, 2005).

Naive Bayes mengasumsikan independensi dari masing-masing feature pada vektor dan untuk setiap vektor tersebut dihitung nilai probabilitas kemunculannya terhadap class. Nilai probabilitas suatu class terhadap feature set merupakan perkalian antara probabilitas kemunculan suatu class dan probabilitas kemunculan masing-masing vektor feature. Proses ini akan diulang untuk setiap kemungkinan class dan text dengan nilai probabilitas maksimum akan terklasifikasi (Younis, 2015).

$$\hat{s} = s \in S \arg \max P(s) \prod_{j=1}^n P(f_j | s)$$

Persamaan (2)

Dimana

$$P(s_i) = \frac{\text{count}(s_i, w_j)}{\text{count}(w_j)}$$

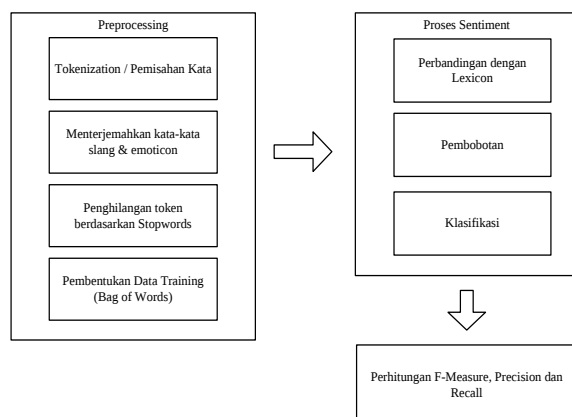
$$P(f_j | s) = \frac{\text{count}(f_j, s)}{\text{count}(s)}$$

Keuntungan penggunaan NBC adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian.

IV. MODEL SISTEM

Berdasarkan model yang telah dirancang, sistem terbagi menjadi 3 tahap, yaitu : tahap preprocessing, proses sentiment, dan perhitungan akurasi berupa F-Measure, Precision, dan Recall. Input terhadap sistem adalah sejumlah teks opini dari Instagram dan outputnya berupa akurasi hasil klasifikasi.

Tahap preprocessing dilakukan untuk mengubah bentuk data dari text menjadi bentuk matriks yang lebih terstruktur sebagai inputan untuk proses klasifikasi. Pada tahap ini text akan diubah menjadi kumpulan kata. Jika terdapat kata-kata slang dan emoticon pada text, maka akan diterjemahkan menjadi suatu kata berdasarkan daftar yang ada. Setelah itu, kata-kata yang mengandung stopwords (kata umum) akan dihapus dari daftar dan kemunculan kata-kata pada setiap text atau komentar akan dihitung untuk membentuk data training yang menjadi inputan untuk tahap berikutnya.



Gambar 4 Model Sistem Sentiment Analysis

Pada tahap proses sentiment, daftar kata akan dibandingkan dengan lexicon untuk mendapatkan nilai pembobotan positif, negatif atau netral. Setelah itu, data training dengan pembobotan akan masuk ke proses klasifikasi untuk penentuan kelas positif, negatif atau netral. Dari proses inilah dapat diketahui tingkat kepercayaan suatu *online shop*.

Pada tahap terakhir, dilakukan perhitungan *F-Measure*, *Precision*, dan *Recall* untuk mengukur tingkat akurasi dari proses klasifikasi.

Penelitian ini akan menggunakan data berupa komentar dari setiap foto dan caption pada 10 buah *online shop* di Instagram dengan jumlah followers minimal 2000 sebagai data training dan dilakukan *Cross Validation* untuk proses testingnya.

V. KESIMPULAN

Sentiment Analysis dapat dilakukan untuk mereview sebuah *online shop* dengan memanfaatkan komentar pada akun Instagram. Model untuk proses review dengan *sentiment analysis* pun telah dirancang dengan menggabungkan pendekatan *lexicon* dan *machine learning*. Model tersebut dapat diterapkan di berbagai bahasa hanya dengan mengganti daftar lexicon dan *slank word*. Dengan menggunakan model tersebut, diharapkan hasil klasifikasi sentimen menghasilkan akurasi yang sangat baik.

Penelitian berikutnya adalah melakukan experiment terhadap model dengan dataset yang bersumber dari Instagram.

REFERENSI

- Das, T. K., Acharjya, D. P., & Patra, M. R. (2014, January). Opinion mining about a product by analyzing public tweets in Twitter. In *Computer Communication and Informatics (ICCCI), 2014 International Conference on* (pp. 1-4). IEEE.
- Fang, Xing, Justin Zhang. (2015). Sentiment Analysis Using Product Review Data. *Journal of Big Data*.
- Ingale, Sonali D. (2015). Sentiment Classification for Product Review Analysis. *International Journal of Engineering Research & Technology (IJERT)* Vol. 4 Issue 07

- Liu, Bing. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publisher
- Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. Expert Systems with Applications, 40(10), 4241-4251.
- Rain, Callen.(2013). Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning. Swarthmore College
- Santos, Antonio Paulo. (2012). A Bootstrapping Algorithm for Learning the Polarity of Words. 10th International Conference, PROPOR 2012, Coimbra, Portugal, April 17-20, 2012.
- Sitaram Asur, Bernardo A. Huberman. (2010). Predicting the Future with Social Media. Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01 Pages 492-499
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307.
- Vinodhini, G., R.M. Chandrasekaran. (2012). Sentiment Analysis and Opinion Mining : A Survey. International Journal of Advanced Research in Computer Science and Software Engineering vol 2, issue 6
- Written ,Ian H., Eibe Frank. (2005). Data Mining : Practical Machine Learning Tools and Techniques 2nd edition. San Francisco : Morgan Kaufmann Publisher
- Younis, Eman MG. (2015). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools : An Empirical Study. International Journal of Computer Application vol 112 – No 5