

Opinion Mining Terhadap Toko Online Di Media Sosial Menggunakan Algoritma Naïve Bayes (Studi Kasus: Akun Facebook Dugal Delivry)

Yustia Hapsari¹, Muhammad Fikri Hidayattullah^{2*}, Dairoh³, Mohammad Khambali⁴

¹Jurusan Teknik Informatika, STMIK YMI Tegal

^{2,3}Jurusan Teknik Informatika, Politeknik Harapan Bersama Tegal

⁴Jurusan Teknik Listrik, Politeknik Negeri Semarang

¹Jl. Pendidikan, Pesurungan Kidul, Kota Tegal, Jawa Tengah 52147, Indonesia

^{2,3}Jln. Mataram No.09 Pesurungan Lor, Kota Tegal 52147, Indonesia

⁴Jln. Prof. H. Soedarto, S.H., Tembalang, Semarang 50275, Indonesia

email: ¹yustia.hapsari@gmail.com, ²fikri@poltektegal.ac.id, ³dairoh@poltektegal.ac.id, ⁴mc.chambali.poltek@gmail.com

Received: 30 Maret 2018; Revised: 12 Mei 2018; Accepted: 13 Mei 2018

Copyright ©2018 Politeknik Harapan Bersama Tegal. All rights reserved

Abstract – The Internet era has had an impact in various sectors of human life. One is the economic sector. Economic transactions change from the traditional pattern (face to face) to online. The customer does not need to ask about the condition of an item to be purchased to a close friend or family, but simply by reviewing the product from the same buyer's comments. Products that get good reviews mean good quality. However, a problem arises if the comment data is very large and will make it difficult for customers to summarize the quality. Therefore, an automatic opinion mining system is required which can directly give conclusions about the quality of a product. This research makes an opinion mining system by applying the Naïve Bayes algorithm by taking a case study of facebook account Dugal Delivry. The measurement result with confusion matrix gives precision value of 88,89%, recall 80% and accuracy equal to 85%.

Abstrak – Era internet memberikan dampak di berbagai sektor kehidupan manusia. Salah satunya adalah sektor ekonomi. Transaksi ekonomi berubah dari pola tradisional (*face to face*) ke *online*. *Customer* tidak perlu bertanya tentang kondisi suatu barang yang akan dibeli ke teman dekat atau keluarganya, namun cukup dengan melihat *review* produk dari komentar para pembeli produk yang sama sebelumnya. Produk yang mendapat *review* bagus berarti berkualitas baik. Namun muncul sebuah masalah jika data komentar tersebut jumlahnya sangat besar dan akan menyulitkan *customer* dalam menimpulkan kualitasnya. Oleh karena itu dibutuhkan sebuah sistem *opinion mining* otomatis yang dapat secara langsung memberikan kesimpulan tentang kualitas suatu produk. Penelitian ini membuat sebuah sistem *opinion mining* dengan menerapkan algoritma *Naïve Bayes* dengan mengambil studi kasus akun facebook Dugal Delivry. Hasil pengujian dengan *confusion matrix* memberikan nilai presisi sebesar 88,89%, *recall* sebesar 80% dan akurasi sebesar 85%.

Kata Kunci : *opinion mining, sentiment analysis, Naïve Bayes*

*) **Corresponding author**: Muhammad Fikri Hidayattullah
Email: fikri@poltektegal.ac.id

I. PENDAHULUAN

Perkembangan teknologi internet membawa dampak perubahan pada beberapa sektor kehidupan manusia, mulai dari pendidikan, perdagangan, pemerintahan hingga komunikasi sosial. Banyak pola baru dalam kehidupan manusia yang muncul setelah adanya perkembangan teknologi internet dan merubah pola lama (tradisional). Salah satu bentuk perubahan pola tersebut yang secara massif terjadi dimana-mana adalah dalam transaksi jual beli. Internet mengubah pola jual beli dari *face to face* menjadi transaksi *online* dengan segala tawaran kemudahannya. Ditambah lagi dengan munculnya media sosial seperti *facebook, twitter, instagram* dan *whatsapp* yang dapat dijadikan sebagai sarana untuk jual beli *online* tanpa perlu membuat toko *online* berbasis *website*.

Salah satu langkah yang biasa dilakukan oleh seorang *customer* sebelum memutuskan untuk membeli barang di toko online adalah melihat *review* dari para pembeli produk yang sama sebelumnya yang menuliskan opini dan tanggapan mereka tentang kondisi barang maupun pelayanan dari toko *online* tersebut. Langkah ini sangat membantu para *customer* sebelum memutuskan membeli sebuah produk tanpa bertanya terlebih dahulu kepada orang lain. Akan tetapi kendala akan muncul manakala jumlah opini dan tanggapan yang ada jumlahnya sangat banyak. Hal ini akan menyulitkan *customer* dalam menarik kesimpulan. Oleh karena itu diperlukan *opinion mining* otomatis yang dapat mengolah kata dan frase tentang ulasan suatu produk untuk membantu memberikan kemudahan dalam menyimpulkan kondisi barang dan pelayanan dari sebuah toko online [1]. *Opinion mining* merupakan nama lain dari *sentiment analysis*. *Opinion mining* merupakan sebuah bidang studi yang menganalisis pendapat, sentimen, emosi, sikap dan penilaian dari orang-orang tentang suatu produk, layanan, individu, organisasi, topik, masalah, peristiwa beserta atributnya [2]. Melalui *opinion mining* dapat diketahui kecenderungan para *customer*

terhadap suatu produk. *Opinion mining* bisa digunakan pada beberapa studi kasus, seperti *review* film [3], mengulas wacana politik pada media *online* [4][5], klasifikasi partai politik [6] hingga dapat digunakan untuk mengukur popularitas kandidat presiden [7][8]. Oleh karena riset di dalam *opinion mining* selalu berkembang terus.

Pada penelitian ini akan melakukan *opinion mining* terhadap komentar-komentar di *facebook* dengan mengambil sampel studi kasus akun *facebook* Dugal Delivry. *Facebook* merupakan salah satu media sosial yang lebih populer di dunia dibandingkan dengan media sosial lainnya. Pada tahun 2012 saja *facebook* sudah memiliki 1 miliar pengguna aktif bulanan dan lebih dari 550 juta pengguna aktif harian [9]. Sedangkan untuk metode klasifikasi menggunakan algoritma *Naïve Bayes*. Algoritma ini memiliki kelebihan yaitu mampu melakukan klasifikasi dengan baik dengan hanya membutuhkan sejumlah kecil *data training* [10]. Hasil dari penelitian ini akan menentukan nilai sentimen positif atau negatif dari kumpulan komentar yang diujikan.

II. PENELITIAN YANG TERKAIT

Penelitian terkait analisis sentimen atau penggalian opini telah dilakukan oleh beberapa peneliti, diantaranya dilakukan oleh Gusriani *et al.*[11] dengan melakukan analisis sentimen terhadap toko online di sosial media menggunakan metode klasifikasi *Naïve Bayes*. Pada penelitian tersebut mengambil objek studi kasus *facebook page* (FP) BerryBenka. Tahapan penelitian yang dilakukan meliputi *data collecting* dengan melakukan *crawling* terhadap komentar-komentar dari para *customer* di FP BerryBenka, dilanjutkan proses *preprocessing* dengan melakukan *punctuation and number filter* dan *case folding*, setelah itu masuk ke tahap *feature selection* dan *attribute selection*, kemudian masuk ke tahap klasifikasi dengan algoritma *naïve bayes*. Metode pengujian menggunakan *k-Fold Cross Validation* dan *Confusion Matrix*. Dari hasil pengujian diperoleh nilai akurasi 93.7%.

Penelitian lain tentang analisis sentimen juga dilakukan oleh Liu *et al.*[12]. Penelitian ini bertujuan untuk mengevaluasi skalabilitas algoritma *Naïve Bayes Classifier* (NBC) ketika digunakan dalam data berskala besar. Sistem analisis *Big Data* juga dirancang untuk penelitian ini. Hasilnya menggembirakan karena akurasi NBC meningkat dan mendekati 82% ketika ukuran dataset juga meningkat.

Pada tahun 2016 Dey *et al.*[10] melakukan penelitian seputar sentimen analisis untuk mendapatkan hasil *review* film dan hotel secara cepat. Pada penelitian ini juga mengkomparasi keakuratan, presisi, serta nilai *recall* dua algoritma klasifikasi untuk kasus sentimen analisis yaitu K-NN dan *Naïve Bayes*. Dan algoritma *Naïve Bayes* memberikan hasil akurasi jauh lebih baik daripada K-NN untuk kasus *review* film yaitu sebesar 82,43%, sedangkan untuk akurasi K-NN-nya sebesar 69,81%. Untuk kasus *review* hotel memberikan hasil akurasi kurang lebih sama, 55,09% untuk *Naïve Bayes* dan 52,14% untuk K-NN.

Penelitian yang khusus berfokus menganalisis performa algoritma *Naïve Bayes* untuk klasifikasi data telah dilakukan oleh Patil dan Sherekar[13]. Pada penelitian ini melakukan studi komparasi antara algoritma *Naïve Bayes* dengan J48 *classification* yang berbasis algoritma *decision tree*. Pada penelitian ini memberikan hasil akurasi klasifikasi untuk

algoritma *Naïve Bayes* dengan klasifikasi YES sebesar 9% dan NO sebesar 89%. Sedangkan untuk algoritma J48 menghasilkan akurasi YES sebesar 31% dan NO sebesar 87%.

III. ALGORITMA NAÏVE BAYES

Algoritma *Naïve Bayes* didasarkan dari aturan Bayes. Meskipun algoritma ini dari sisi penamaan kurang meyakinkan karena menggunakan istilah “*naïve*”, namun memiliki performa yang bagus ketika diuji cobakan pada *dataset* yang aktual [14]. Di dalam kasus *opinion mining* algoritma ini digunakan untuk mengklasifikasikan hasil sentimen.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

$P(C_i|X)$: probabilitas posterior

$P(X|C_i)$: probabilitas prior

$P(X|C_i)$: *likelihood*

$P(X)$: *evidence*

Atau dapat juga dituliskan [15]:

$$Posterior = \frac{Likelihood * Prior Probability}{Evidence} \quad (2)$$

IV. METODE PENELITIAN

A. Alur Perancangan

Di dalam penelitian ini tahapan awal dari pengumpulan data hingga pengujian akurasi *opinion mining* dapat dilihat di Gbr 1.

a) Data Collecting

Tahap awal dari penelitian ini adalah pengumpulan data (*data collecting*). Pada tahap ini melakukan pengambilan data-data komentar dari akun *facebook* Dugal Delivry. Setelah data terkumpul akan dilanjutkan dengan proses seleksi *data training* (*training data selection*).

b) Preprocessing

Tahapan *preprocessing* merupakan tahapan persiapan awal sebelum masuk ke proses berikutnya [15]. Di dalam tahap ini ada tiga proses yang dilakukan. Pertama adalah *Tokenization*. Tahap ini berguna untuk memecah data-data komentar dalam bentuk kalimat yang akan diproses. Di dalam tahap ini juga terjadi penghapusan tanda baca dan angka. Tahap kedua adalah *Stop Word Removal*. Tahap ini berfungsi menghilangkan kata-kata yang sering muncul dan tidak begitu memiliki makna dalam proses *opinion mining*, seperti kata “dan”, “yang”, “atau” dan lain sebagainya. Sedangkan tahap ketiga yaitu *stemming*, berfungsi menjadikan kata yang telah mendapatkan sisipan kembali ke kata dasarnya, misalkan “membelikan” akan berubah menjadi “beli”.

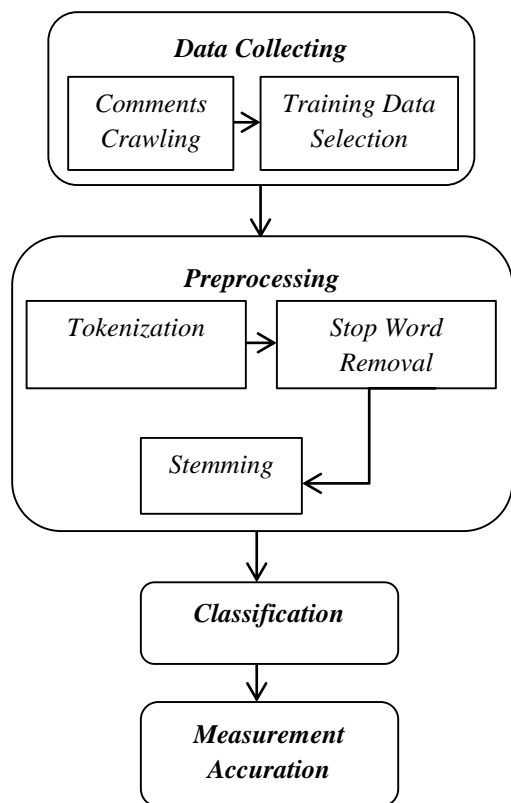
c) Classification

Classification merupakan tahap terakhir dalam *opinion mining*. Tahap ini berfungsi untuk menentukan nilai sentimen terhadap suatu produk atau layanan. Algoritma yang

digunakan dalam tahap ini adalah *Naïve Bayes*. Di dalam penelitian ini hanya ada dua nilai hasil sentimen yaitu positif dan negatif.

d) *Measurement Accuration*

Tahap terakhir ini bertujuan untuk menguji tingkat akurasi *opinion mining* dengan algoritma *Naïve Bayes*. Metode perhitungan akurasi yang dilakukan dengan menggunakan *confusion matrix*.



Gbr. 1 Tahapan *Opinion Mining*

V. HASIL DAN PEMBAHASAN

Data training yang digunakan dalam penelitian ini berjumlah sebanyak 49 komentar bernilai positif dan 26 komentar bernilai negatif. Kumpulan komentar negatif dan positif yang dijadikan sebagai *data training* dibutuhkan untuk penentuan hasil pengujian *data testing*. Komentar-komentar tersebut diambil dari data komentar di akun *facebook* Dugal Delivry. Sedangkan untuk *data testing* berjumlah 10 data bernilai positif dan 10 data bernilai negatif. *Data testing* digunakan untuk menghitung nilai akurasi. *Tools* yang digunakan untuk penentuan nilai opini adalah Rapid Miner.

Setelah seluruh *data training* terkumpul semua akan dilanjutkan tahap *preprocessing* dengan melakukan proses penghilangan tanda baca dan angka yang tertulis di komentar-komentar para *customer*. Tujuan dari *preprocessing* ini adalah untuk membersihkan *data training* dari *noise*, mengurangi volume jumlah kata dan menyeragamkannya. Tanpa melalui proses *preprocessing*

akan ditemukan banyak data yang mengandung kata-kata yang kurang berarti dan memiliki makna.

Tahap selanjutnya adalah *feature selection*. Tahap ini berisi *filtering* dan *stemming*. *Stemming* merupakan suatu proses mengubah sebuah kata yang telah mendapatkan sisipan menjadi kata dasar. Hanya saja kamus standar *stemming* Bahasa Indonesia di Rapid Miner belum tersedia, sehingga harus memanfaatkan kamus eksternal. Tahapan *feature selection* ini sangat diperlukan sebelum masuk ke proses pelatihan data di *Naïve Bayes*.

Tahap terakhir sebelum masuk pengujian akurasi dalam *opinion mining* adalah tahap klasifikasi. Algoritma klasifikasi yang digunakan adalah *Naïve Bayes*. Tahap ini akan menghasilkan klasifikasi opini positif atau negatif.

TABEL I
HASIL PENGUJIAN DATA TESTING

No.	Kalimat yang dianalisis	Nilai opini (sentimen)
1.	ah mimin banyak diemnya kalo dikomen..	Negatif
2.	Kambing lo...	Negatif
3.	Lama bgt datangnya	Negatif
4.	Min, klapertartnya keenceran nih	Negatif
5.	bisa lbh cepet ga dtgnya...	Negatif
6.	Dikomen ga balas-balas	Negatif
7.	Payah nih mimin..	Negatif
8.	rasanya tidak enak	Positif
9.	Jangan mahal-mahal dong.. kasih diskon banyakin	Negatif
10.	Lama nunggunya om admin	Negatif
11.	Lezat banget kuenya	Positif
12.	Mantab deh pokoknya..duriannya kerasa banget	Positif
13.	Enak..	Positif
14.	bikin anaku ketagihan min.. yummy	Positif
15.	wah, pengen nambah lagi pesenannya.. enak sih soalnya	Positif
16.	Yummi..	Negatif
17.	Alhamdulillah, pesenanku dah sampai min.. mantab deh pokoknya	Positif
18.	rasanya mantab dan lezat...jos	Negatif
19.	Rasanya emang enak bgt	Positif
20.	Alhamdulillah dah nyampe mbak..	Positif

Kesalahan hasil klasifikasi pada *data testing* negatif terdapat pada kalimat “rasanya tidak enak”. Kesalahan klasifikasi ini disebabkan tidak mempunya sistem *opinion mining* ini menegasikan kata “tidak” (*convert negation*). Sehingga setiap kali ada kata “tidak” disambungkan dengan kata bermuatan opini positif akan dianggap sebagai opini positif. Padahal kalimat tersebut merupakan opini negatif dan bentuk negasi dari opini positif.

Pengujian akurasi menggunakan *confusion matrix* yang meliputi perhitungan presisi, *recall* dan akurasi. Proses perhitungan ini berdasarkan perolehan jumlah *True Positive*

(TP), True Negative (TN), False Positive (FP) dan False Negative (FN).

TABEL II
KLASIFIKASI CONFUSION MATRIX

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

Dari hasil pengujian diperoleh nilai:

$$\text{Presisi} = \frac{TP}{FP + TP} * 100\% \quad \text{Recall} = \frac{TP}{FN + TP} * 100\%$$

$$\text{Presisi} = \frac{8}{1+8} * 100\% = 88,89\% \quad \text{Recall} = \frac{8}{2+8} * 100\% = 80\%$$

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

$$\text{Akurasi} = \frac{8+9}{8+9+1+2} * 100\% = 85\%$$

Melalui pengujian *confusion matrix* diperoleh hasil untuk presisi sebesar 88,89%, *recall* sebesar 80% dan akurasi sebesar 85%.

VI. KESIMPULAN

Algoritma *Naïve Bayes* terbukti cukup handal diterapkan dalam kasus klasifikasi opinion mining. Algoritma ini mampu menghasilkan nilai presisi sebesar 88,89%, *recall* sebesar 80% dan akurasi sebesar 85%. Namun, jika diamati salah satu penyebab kesalahan klasifikasi adalah tidak adanya fitur *convert negation*. Fitur *convert negation* sangat diperlukan sekali karena untuk menghindari kesalahpahaman *classifier* ketika mengklasifikasikan frase atau kalimat yang mengandung kata “tidak”. Oleh karena itu untuk pengembangan ke depan perlu ditambah fitur *convert negation* dan perlu adanya optimasi algoritma *Naïve Bayes*.

DAFTAR PUSTAKA

[1] R. Othman, R. Belkaroui, and R. Faiz, “Extracting Product Features Features for for Opinion Opinion Mining Mining

Using Public Conversations in Twitter Using Public Conversations in Twitter,” *Procedia Computer Science*, vol. 112, pp. 927–935, 2017.

[2] B. Liu, “Sentiment Analysis and Opinion Mining,” *Synthesis Lectures on Human Language Technologies*, no. May, 2012.

[3] L. L. Dhande and P. G. K. Patnaik, “Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier,” *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 3, no. 4, pp. 313–320, 2014.

[4] M. Ringsquandl and D. Petković, “Analyzing Political Sentiment on Twitter,” *Analyzing Microtext: Papers from the 2013 AAAI Spring Symposium*, pp. 40–47, 2013.

[5] A. Bakliwal, J. Foster, J. Van Der Puil, R. O. Brien, L. Tounsi, and M. Hughes, “Sentiment Analysis of Political Tweets: Towards an Accurate Classifier,” *Proceedings of the Workshop on Language in Social Media (LASM 2013)*, no. LasM, pp. 49–58, 2013.

[6] T. Ding, J. Deng, J. Li, and Y. Lin, “Sentiment Analysis and Political Party Classification in 2016 U . S . President Debates in Twitter,” *International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation*, pp. 1–6, 2016.

[7] R. Malinský and I. Jelinek, “Sentiment Analysis: Popularity of Candidates for the President of the United States,” *International Journal of Humanities and Social Sciences*, vol. 6, no. 12, pp. 3679–3681, 2012.

[8] E. Tunggawan, “And the Winner is ...: Bayesian Twitter-based Prediction on 2016 U . S . Presidential Election,” *IC3INA 2016 Conference (3-5, no. 1, pp. 1–5, 2016*.

[9] A. Ortigosa, J. M. Martín, and R. M. Carro, “Computers in Human Behavior Sentiment analysis in Facebook and its application to e-learning,” *COMPUTERS IN HUMAN BEHAVIOR*, 2013.

[10] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, “Sentiment Analysis of Review Datasets using Naïve Bayes ’ and K-NN Classifier,” *Information Retrieval (cs.IR); Computation and Language (cs.CL)*, vol. 8, no. 4, pp. 54–62, 2016.

[11] S. Gusriani, K. D. K. Wardani, and M. I. Zul, “Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus ... Analisis Sentimen Terhadap Toko Online di Sosial Media,” *4th Applied Business and Engineering Conference*, no. November, 2016.

[12] B. Liu, E. Blasch, Y. Chen, D. Shen, E. Blasch, Y. Chen, D. Shen, and G. Chen, “Scalable sentiment classification for Big Data analysis using Naïve Bayes Classifier Scalable Sentiment Classification for Big Data Analysis Using Naïve Bayes Classifier,” *IEEE Intl Conf. on Big Data*, no. October 2013, 2014.

[13] T. R. Patil and S. S. Sherekar, “Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification,” *International Journal Of Computer Science And Applications*, vol. 6, no. 2, 2013.

[14] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Third Edition*. Elsevier Inc., 2011.

[15] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, 12th ed. Springer, 2011.