

Klasifikasi Tweet E-Commerce dengan Menggunakan Metode Support Vector Machine

Al Hafiz Yunas¹, Yusra², Muhammad Fikry³

^{1,2,3}Teknik Informatika, UIN Sultan Syarif Kasim Riau

Jl. H.R. Soebrantas no. 155 KM. 18 Simpang Baru, Pekanbaru 28293

al.hafiz.yunas@students.uin-suska.ac.id¹, yusra@uin-suska.ac.id², muhammad.fikry@uin-suska.ac.id³

Abstrak – Aktifitas belanja *online* telah menjadi kebutuhan masyarakat. *Online shop* di media sosial merupakan pilihan tempat berbelanja karena pembeli dapat berinteraksi dan berkonsultasi langsung dengan penjual. Tantangan dalam mengumpulkan informasi transaksi *e-commerce* di media sosial adalah banyaknya pemilik *online shop* dan kerahasiaan data. Namun demikian, informasi transaksi *e-commerce* di Twitter dapat ditemukan pada *tweet* yang dapat diakses publik. *Tweet* biasanya berisikan aktifitas sebelum pembelian, aktifitas pembelian, aktifitas pengiriman oleh penjual, atau aktifitas penerimaan oleh pembeli. Hal ini menjadi indikator adanya transaksi. Tantangan lainnya adalah teks di media sosial menggunakan bahasa alami manusia yang seringkali dituliskan secara tidak terstruktur. Dalam penelitian ini, diklasifikasikan apakah suatu *tweet* berkaitan dengan transaksi *e-commerce* atau tidak. Oleh karena itu, *tweet* yang telah dikumpulkan dan diberi label perlu dipraproses, meliputi *case folding*, *cleaning*, tokenisasi, normalisasi kata, *stopword removal* dan *stemming*. Selanjutnya, dilakukan ekstraksi fitur berdasarkan nilai *document frequency* dan *threshold minimum* bagi kata untuk dipilih sebagai fitur. Untuk setiap *tweet*, ditentukan nilai fitur dengan *term frequency-inverse document frequency*. Setelah dilakukan *cross-validation* dengan menggunakan *kernel RBF*, diketahui parameter terbaik adalah pasangan parameter $C=0,9$ dan $\gamma=0,8$ dengan rataan akurasi sebesar 96,1%. Model terbaik merupakan model yang menghasilkan nilai akurasi tertinggi. Akhirnya dilakukan pengujian dengan hasil akurasi sebesar 94%.

Kata Kunci – E-Commerce, Klasifikasi, SVM, Transaksi Online, Tweet

PENDAHULUAN

Seiring perkembangan zaman dan kemajuan teknologi informasi, aktifitas belanja *online* tidak hanya menjelma menjadi gaya hidup bagi banyak orang, bahkan telah menjadi suatu kebutuhan. Barang yang diinginkan dapat dipesan melalui *internet*, dibayar melalui transfer ke bank atau menggunakan kartu kredit, kemudian barang tersebut diantar sampai ke depan rumah. Aktifitas menjual

atau membeli barang secara *online* ini disebut dengan *electronic commerce* (disingkat *e-commerce*).

Ada banyak manfaat yang diperoleh penjual maupun pembeli melalui *e-commerce*. Penjual dapat menjual barangnya secara global, mengurangi infrastruktur yang diperlukan, mengurangi biaya operasional, sehingga harga barang menjadi lebih murah. Pembeli dapat berbelanja kapan saja dan dimana saja, memiliki banyak pilihan dan dapat membandingkan harganya.

Bentuk bisnis *e-commerce* yang umum dijumpai di Indonesia adalah *Business-to-Customer* (B2C) berupa *online shop* (baik di situs *web* maupun di media sosial) dan *Customer-to-Customer* (C2C) berupa *marketplace*. Perilaku berbelanja secara *online* ini terlihat dari hasil survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2016 [1]. Pengguna *internet* Indonesia pada tahun 2016 berjumlah 132,7 juta orang atau 51,8% dari total populasi penduduk Indonesia. Sebanyak 98,6% responden mengetahui *internet* sebagai tempat jual beli barang dan jasa, sementara sebanyak 63,5% responden pernah bertransaksi secara *online*. *Online shop* merupakan konten komersial yang paling sering dikunjungi oleh pengguna *internet* Indonesia, sebanyak 62% responden.

Online shop di media sosial menjadi pilihan tempat berbelanja karena calon pembeli dapat berinteraksi dan berkonsultasi langsung dengan penjual. Sebagai salah satu media sosial dengan jumlah pengguna terbesar dunia, Twitter sangat potensial dalam perkembangan *e-commerce*. Banyaknya jumlah *followers* akun *e-commerce* baik perusahaan maupun perorangan serta banyaknya *tweet* yang berhubungan dengan aktifitas *e-commerce*, menjadi alasan bagi pemilik Twitter untuk menaruh perhatian terhadap penggunaan Twitter untuk *e-commerce* [2]. Perhatian tersebut diperlihatkan dengan penambahan fitur “*buy*” pada *tweet* yang berkarakteristik *e-commerce*.

Pemerintah Indonesia juga menaruh perhatian terhadap transaksi terkait *e-commerce*. Peningkatan jumlah transaksi yang tajam telah mendorong pemerintah untuk menggali potensi pajak pada transaksi jual beli dan iklan *e-commerce* termasuk di media sosial yang selama ini belum tersentuh [3]. Pertumbuhan bisnis *e-commerce* diharapkan menjadi sumber penerimaan negara sebagaimana bisnis konvensional.

Informasi transaksi *e-commerce* di media sosial sulit untuk dikumpulkan dikarenakan beberapa hal. Pertama, banyaknya pemilik *online shop*. Berdasarkan

hasil survei APJII tahun 2017 [4], Pengguna *internet* Indonesia pada tahun 2017 berjumlah 143,26 juta orang atau 54,68% dari total populasi penduduk Indonesia. Sebanyak 16,83% responden menggunakan *internet* untuk berjualan secara *online*. Kedua, data transaksi bersifat rahasia bagi pemilik *online shop*. Terlebih lagi, sistem pemungutan pajak di Indonesia menganut sistem *self-assessment* dimana wajib pajak yang menghitung, membayar dan melaporkan sendiri jumlah pajak yang terutang.

Namun demikian, informasi transaksi *e-commerce* di media sosial dapat ditemukan pada teks (misalnya *tweet*) yang dapat diakses publik. Teks biasanya berisikan aktifitas sebelum pembelian, aktifitas pembelian, aktifitas pengiriman oleh penjual, atau aktifitas penerimaan oleh pembeli. Hal ini dapat menjadi indikator adanya transaksi.

Saat ini, masih minim penelitian terkait informasi transaksi *e-commerce* di *tweet* berbahasa Indonesia. Penelitian pertama kali mengenai subyek tersebut adalah klasifikasi transaksi *e-commerce* melalui pengembangan aplikasi SAFE-F [5]. Dalam penelitian tersebut, dilakukan klasifikasi *tweet* dan ekstraksi 10 jenis informasi. Klasifikasi *tweet* dengan menggunakan algoritma pembelajaran C4.5, fitur trigram dan tanpa praproses menunjukkan akurasi 85,0%. Ekstraksi informasi dengan menggunakan algoritma *instance-based learning* menunjukkan akurasi 81,49%. Penelitian berikutnya berfokus kepada ekstraksi 2 jenis informasi yaitu lokasi dan produk berbasis pada model klasifikasi Naïve Bayes [6]. Hasil pengujian menunjukkan nilai *precision* mencapai 96% dan *recall* mencapai 83%.

Salah satu metode pembelajaran yang lebih baru dibandingkan dengan C4.5 dan Naïve Bayes adalah Support Vector Machine (SVM). Beberapa penelitian yang menggunakan metode SVM untuk mengklasifikasikan *tweet* sebagai berikut. Penelitian berjudul 'Penerapan Metode SVM Menggunakan Kernel Radial Basis Function (RBF) pada Klasifikasi Tweet' untuk mengklasifikasikan iklan [7]. Penelitian tersebut menunjukkan akurasi sebesar 99,12% dengan melakukan pemilihan fitur, dan akurasi 97,54% tanpa memilih fitur. Penelitian lainnya berjudul 'Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik' [8]. Penelitian tersebut menunjukkan metode Support Vector Machine menghasilkan akurasi performansi yang lebih baik daripada metode Naïve Bayes, yaitu 83,14% berbanding 73,81%.

Pada penelitian ini, digunakan metode SVM untuk mengklasifikasikan apakah suatu *tweet* berkaitan dengan transaksi *e-commerce* atau tidak. Hasil klasifikasi akan dievaluasi untuk mengetahui tingkat akurasinya. Hal ini merupakan langkah awal untuk memahami informasi transaksi *e-commerce* di media sosial. Kedepannya, penelitian ini akan dilanjutkan ke langkah berikutnya yaitu mengklasifikasikan aktifitas pada *tweet* terkait transaksi *e-commerce*, dan mengekstraksi jenis informasi yang ada.

LANDASAN TEORI

A. Tweet

Tweet adalah status yang dibuat di media sosial Twitter dalam 280 karakter atau kurang. Sebelumnya batasan jumlah karakter adalah 140 karakter. Teks *tweet* dapat berisikan URL, *mention*, dan *hashtag*. *Tweet* dapat diunduh melalui Twitter API dalam format JavaScript Object Notation (JSON). Obyek *tweet* terdiri atas atribut (misalnya *id*, *created_at* dan *text*) dan obyek lain (misalnya *user*, *entities* dan *extended_entities*). Obyek *user* berisikan metadada akun Twitter. Obyek *entities* berisikan metadada dan informasi tambahan, seperti URL, *mention*, *hashtag*, dan *symbol*. Obyek *extended_entities* berisikan metadada untuk media meliputi foto, video, dan animated GIF [9].

B. E-Commerce

E-commerce didefinisikan sebagai proses pembelian, penjualan, mentransfer atau bertukar produk, jasa atau informasi melalui *internet* [10]. *E-commerce* juga dapat didefinisikan sebagai transaksi digital yang terjadi melalui *internet*, *web*, dan/atau aplikasi bergerak (*mobile application*) [11].

Mekanisme transaksi *e-commerce* yaitu [12]

1. Pembeli melihat produk atau jasa yang diiklankan. Pembeli juga mencari data atau informasi tertentu yang dibutuhkan terkait transaksi yang akan dilakukan.
2. Jika tertarik, pembeli dapat memesan dengan dua cara, yaitu :
 - a. Secara konvensional melalui telepon, faks, atau datang langsung ke tempat penjualan.
 - b. Secara elektronik dengan menggunakan perangkat komputer.
3. Berdasarkan pesanan konsumen, penjual akan mendistribusikan barangnya kepada pembeli melalui dua cara, yaitu :
 - a. Untuk produk fisik, dikirimkan melalui kurir ke tempat konsumen.
 - b. Untuk produk digital, dikirimkan melalui *internet*.
4. Aktifitas pasca transaksi berupa pelayanan purna jual (*electronic customer support*). Proses ini dapat dilakukan melalui jalur konvensional atau *internet*.

C. Support Vector Machine

Secara umum, permasalahan klasifikasi berkaitan dengan mempelajari apa yang memisahkan misalnya dua kumpulan obyek dan berdasarkan hal tersebut meletakkan suatu obyek yang sebelumnya tidak diketahui ke dalam salah satu kumpulan. SVM adalah suatu metode untuk menemukan *hyperplane* yang dapat memisahkan dua kumpulan data dari dua kelas yang berbeda [13] dengan *margin* terbesar. *Margin* merupakan jarak antara *hyperplane* dengan data terdekat pada masing-masing kelas. Data-data terdekat dengan *hyperplane* inilah yang disebut *support vector*.

Untuk melakukan klasifikasi, data dipisahkan menjadi kumpulan data latih (*training*) dan kumpulan data

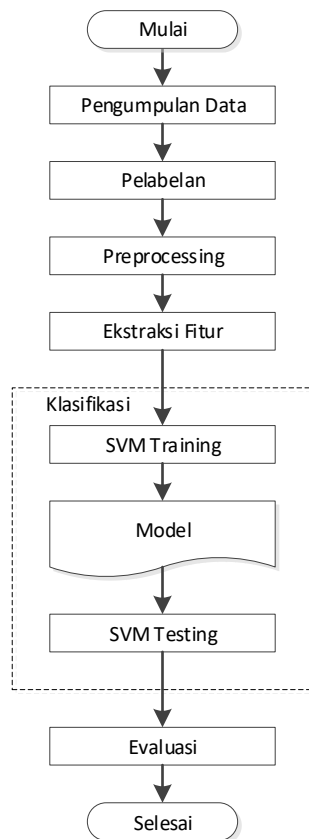
uji (*testing*). Setiap data mengandung satu nilai target yaitu label kelas, dan sejumlah atribut yaitu ciri (*feature*) atau variabel yang diamati. Tujuan SVM adalah untuk menghasilkan suatu model (berdasarkan data latih) yang dapat memprediksi label kelas dari data uji berdasarkan atribut data uji [14].

Hyperplane tidak selalu berbentuk garis lurus pada dua dimensi ataupun bidang datar pada tiga dimensi. Metode SVM dapat bekerja pada data linier maupun non linier. Pada permasalahan klasifikasi dengan kasus yang tidak dipisahkan secara linier, dipergunakan fungsi *kernel*, juga dikenal sebagai *kernel trick* [15]. *Kernel* dapat memetakan dimensi pada ruang masalah menjadi dimensi yang lebih tinggi dimana data-data tersebut mungkin dapat dipisahkan secara linier.

Terdapat sejumlah *kernel* yang umum digunakan, yaitu *linear kernel*, *gaussian radial basis function (RBF) kernel*, *exponential kernel*, *polynomial kernel*, *hybrid kernel*, dan *sigmoidal*. Pada penelitian ini, kernel yang digunakan adalah *kernel RBF* dengan pasangan parameter C dan γ .

METODE PENELITIAN

Tahapan-tahapan yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Dataset yang digunakan adalah *tweet* yang mengandung informasi terkait transaksi *e-commerce* dan *tweet* yang tidak mengandung informasi tersebut misalnya iklan. *Tweet* diunduh melalui Twitter API dari sejumlah akun *e-commerce*, kemudian disimpan ke dalam basis data. Jumlah *tweet* sebanyak 1000 dengan rincian per kelas diperlihatkan pada Tabel 1, dan pemisahnya menjadi kumpulan data latih dan uji diperlihatkan pada Tabel 2.

Tabel 1. Rincian *dataset* per kelas

Kelas <i>Tweet</i>	Jumlah
Transaksi	500
Non Transaksi	500
Jumlah	1000

Tabel 2. Pembagian *dataset*

Pembagian <i>Tweet</i>	Kelas <i>Tweet</i>	Jumlah
Data latih (80% <i>dataset</i>)	Transaksi	400
	Non Transaksi	400
Jumlah data latih		800
Data latih (20% <i>dataset</i>)	Transaksi	100
	Non Transaksi	100
Jumlah data uji		200

B. Pelabelan Data

Setiap *tweet* diberikan label kelas secara manual. Kelas-kelas yang digunakan yaitu :

1. Kelas Transaksi

Merupakan *tweet* yang mengandung informasi sedang atau telah terjadinya transaksi *e-commerce*. Aktifitasnya meliputi pemesanan, pembayaran, pengiriman, dan penerimaan barang. Pembatalan transaksi dan pengalaman pengguna setelah bertransaksi atau menerima barang juga termasuk ke dalam kelas ini.

Contoh *tweet* sebagai berikut.

Thanks Bandtemenloe kirimannya sudah sampai dan Packaging ok Kali kedua saya langganan sama anda Hasilnya memu

2. Kelas Non Transaksi

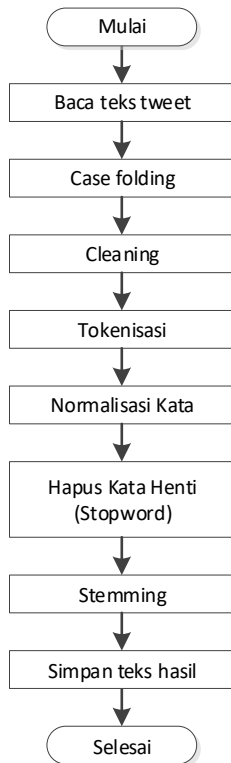
Merupakan *tweet* yang tidak mengandung informasi transaksi *e-commerce*, misalnya pertanyaan sebelum membeli, kuis, berita, dan iklan.

Contoh *tweet* sebagai berikut.

Perlengkapan rumah tangga dari kamar mandi kamar tidur isi dapur hingga furniture bisa cicilan 0%

C. Preprocessing

Dokumen teks yang bersumber dari media sosial merupakan data yang belum terstruktur. Untuk mengubahnya menjadi data yang terstruktur dan dapat digunakan pada proses klasifikasi, dilakukan *preprocessing* terlebih dahulu. Langkah-langkahnya diperlihatkan pada Gambar 2.



Gambar 2. *Preprocessing* data

Detail langkah-langkah *preprocessing* terhadap setiap dokumen teks (*tweet*) yang dibaca dari basis data sebagai berikut.

1. Langkah *case folding* digunakan untuk menyeragamkan semua huruf dalam teks menjadi huruf kecil atau huruf besar. Strategi umum yang digunakan adalah mengubah menjadi huruf kecil.
2. Langkah *cleaning* digunakan untuk membersihkan *noise* dari teks. Dalam penelitian ini, pembersihan dilakukan dengan mengubah *noise* menjadi karakter spasi. Entitas *tweet* yang dibersihkan meliputi URL, *mention* dan *hashtag* yang dikenali dengan ciri berikut.
 - a. URL diawali 'http', 'https', 'ftp' atau 'file'.
 - b. *Mention* diawali simbol '@' yang menunjukkan sebuah akun Twitter.
 - c. *Hashtag* diawali karakter '#'.

Sementara itu, karakter yang dibersihkan meliputi karakter HTML, *emoticon*, angka dan tanda baca.

3. Langkah tokenisasi digunakan untuk memisahkan teks menjadi token-token yang bermakna. Dalam penelitian ini, teks dipisahkan menjadi kata berdasarkan karakter pemisahannya yaitu spasi.

4. Langkah normalisasi kata digunakan untuk mengubah kata tak baku yang ditemukan di dalam teks menjadi kata baku. Pendekatan yang digunakan berbasis kamus (*dictionary-based*), dimana disediakan kamus normalisasi berisikan daftar kata tidak baku dan kata penggantinya.
5. Langkah hapus kata henti (*stopword removal*) digunakan untuk menghapus kata yang tidak penting, tidak bermakna. Kata yang dihapus adalah kata hubung (konjungsi).
6. Langkah *stemming* digunakan untuk mengubah kata ke dalam bentuk dasarnya (*root*). Dalam penelitian ini, digunakan algoritma *stemming* Enhanced Confix Stripping (ECS).

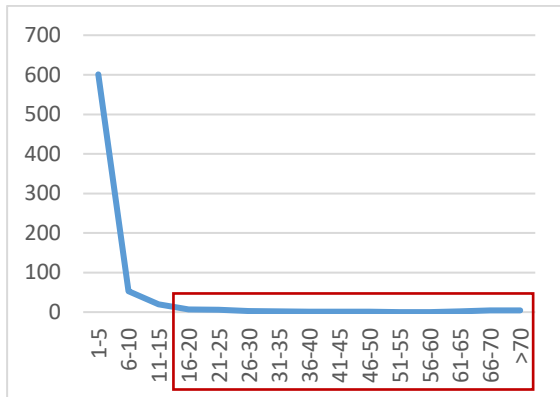
D. Ekstraksi Fitur

Fitur (ciri) yang digunakan dalam klasifikasi ini adalah kata, dengan jenis fitur unigram. Tidak semua kata hasil *preprocessing* yang akan digunakan. Oleh sebab itu, dilakukan penilaian pentingnya setiap kata dengan menggunakan Document Frequency (DF). Tabel 3 memperlihatkan perangkingan kata untuk mengetahui distribusinya.

Tabel 3. Distribusi nilai DF

No	Nilai DF	Jumlah Kata
1	1-5	601
2	6-10	53
3	11-15	20
4	16-20	7
5	21-25	6
6	26-30	3
7	31-35	2
8	36-40	1
9	41-45	1
10	46-50	1
11	51-55	0
12	56-60	0
13	61-65	2
14	66-70	4
15	>70	4

Selanjutnya, ditentukan nilai *threshold* minimum bagi kata untuk dipilih sebagai fitur. Nilai tersebut dipilih berdasarkan konstannya jumlah kata pada grafik sebagaimana diperlihatkan pada Gambar 3. Jumlah kata terlihat mulai konstan pada titik antara 16-20. Nilai *threshold* yang dipilih adalah 16.



Gambar 3. Distribusi nilai DF untuk fitur terpilih

E. SVM Training

Untuk setiap dokumen teks (*tweet*), ditentukan nilai setiap fitur (*vector*) terpilih tersebut dengan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Ketika menghitung nilai TF-IDF untuk data latih dan data uji, total jumlah dokumen yang digunakan mengacu kepada jumlah data latih.

Pada penelitian ini, digunakan salah satu implementasi dari metode SVM yaitu LibSVM [16]. Pada tahapan SVM *training*, dilakukan *10-fold cross validation* terhadap data latih dengan menggunakan *kernel RBF* dan sejumlah pasangan parameter *C* dan γ . Adapun rentang yang digunakan untuk pasangan parameter *C* dan γ yaitu $0.1 \leq C \leq 1$ dan $0.1 \leq \gamma \leq 1$.

Model dengan akurasi terbaik dari nilai pasangan parameter *C* dan γ terbaik akan digunakan terhadap data uji pada tahapan SVM *testing*. Nilai akurasi dapat dirumuskan sebagai berikut :

$$Akurasi = \frac{\text{Jumlah klasifikasi benar}}{\text{Jumlah dokumen}} \times 100\% \quad (1)$$

Untuk mengetahui pasangan parameter terbaik, dilakukan *grid search*. Tabel 4 tersebut berisikan rata-rata akurasi dari pasangan parameter yang rentang nilainya telah ditentukan.

Tabel 4. Rataan akurasi dari pasangan parameter

$\gamma \backslash C$	0.1	0.2	0.3	0.4	0.5
0.1	89,1%	90,7%	89,9%	89,2%	89,5%
0.2	92,4%	92,5%	92,8%	91,8%	90,6%
0.3	91,8%	92,2%	92,9%	93,5%	92,9%
0.4	92,2%	92,7%	93,5%	94,1%	94,0%
0.5	92,0%	92,6%	92,7%	94,2%	94,5%
0.6	92,2%	92,3%	93,0%	94,4%	95,1%
0.7	92,5%	92,2%	93,2%	95%	95,5%
0.8	92,1%	92,6%	93,4%	95,7%	95,8%
0.9	91,9%	92,7%	93,6%	95,9%	95,8%
1.0	91,9%	92,9%	93,8%	95,9%	95,8%

$\gamma \backslash C$	0.6	0.7	0.8	0.9	1,0
0.1	89,2%	89,2%	89,2%	87,2%	87,2%
0.2	89,9%	89,3%	89,3%	87,2%	87,2%
0.3	91,6%	90,8%	90,9%	89,6%	87,5%
0.4	93,1%	91,7%	91,5%	91,1%	90,6%
0.5	93,8%	93,1%	92,7%	91,6%	90,7%
0.6	94,6%	94,2%	93,0%	92,7%	92,2%
0.7	95,5%	95,6%	94,5%	93,5%	92,8%
0.8	95,8%	95,8%	95,6%	95,2%	93,8%
0.9	95,7%	95,8%	96,1%	95,4%	95,3%
1.0	95,5%	95,6%	96,0%	96,0%	95,9%

Berdasarkan Tabel 4, diketahui parameter terbaik adalah pasangan parameter $C=0,9$ dan $\gamma=0,8$ dengan rata-rata akurasi sebesar 96,1%. Model terbaik merupakan model yang menghasilkan nilai akurasi tertinggi.

F. SVM Testing

Pada tahapan SVM *testing*, dilakukan pengujian dengan menggunakan model terbaik yang diperoleh sebelumnya. Data uji terdiri atas 200 *tweet*, meliputi 100 *tweet* transaksi dan 100 *tweet* non transaksi. Dengan menerapkan *threshold*, dilakukan pengujian terhadap 138 *tweet* yang mengandung fitur. Dari hasil pengujian, diperoleh hasil akurasi sebesar 94%.

KESIMPULAN DAN SARAN

Penelitian ini menerapkan metode SVM dalam mengklasifikasi *tweet* terkait transaksi *e-commerce* dan non transaksi *e-commerce*. Metode SVM berhasil menghasilkan model pembelajaran yang mempunyai nilai akurasi klasifikasi sebesar 96%, dengan pasangan parameter $C=0,9$ dan $\gamma=0,8$. Setelah dilakukan pengujian, diperoleh hasil akurasi sebesar 94%. Dari 138 data uji yang mengandung fitur setelah *threshold*, terdapat 8 data uji yang salah kelas. Hal ini terjadi karena fitur yang digunakan mewakili kedua kelas yang diklasifikasi.

Kedepannya, perlu diklasifikasikan *tweet* terkait transaksi *e-commerce* berdasarkan jenis aktifitasnya meliputi pemesanan, pembayaran, pengiriman, penerimaan barang, pembatalan, dan pengalaman pengguna. Selain itu, perlu dilakukan ekstraksi jenis informasi yang ada di dalam *tweet* meliputi nama produk, harga produk, jumlah produk, cara pembayaran, jumlah pembayaran, kepuasan pelanggan, dan harapan pelanggan.

REFERENSI

- [1] Asosiasi Penyelenggara Jasa Internet Indonesia, "Infografis Penetrasi dan Perilaku Pengguna Internet Indonesia," November 2016. [Online]. Available: <https://apjii.or.id/download/file/surveipenetrasiinternet2016.pdf>.
- [2] E. Anggraini, "Twitter Rambah Bisnis E-Commerce?," Februari 2014. [Online]. Available: <http://tekno.liputan6.com/read/817032/twitter-rambah-bisnis-ie-commerce>. [Diakses Januari 2018].
- [3] M. Reily, "BPS dan Pelaku E-commerce Akan Hitung Transaksi Online di Media Sosial," Desember 2017. [Online]. Available: <https://katadata.co.id/berita/2017/12/08/bps-dan-pelaku-e-commerce-akan-hitung-transaksi-online-di-media-sosial>. [Diakses Januari 2018].
- [4] Asosiasi Penyelenggara Jasa Internet Indonesia, "Infografis Penetrasi dan Perilaku Pengguna Internet Indonesia," 2017. [Online]. Available: <https://apjii.or.id/survei2017>.
- [5] M. L. Kodra dan A. Purwarianti, "Ekstraksi Informasi Transaksi Online Pada Twitter," *Jurnal Cybermatika*, vol. 1, no. 1, 2013.
- [6] L. N. Wulansari, "Ekstraksi Lokasi dan Produk dari Data Transaksi Online pada Twitter," 2015.
- [7] I. A. Muis dan M. Affandes, "Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet," *Jurnal Sains, Teknologi dan Industri*, vol. 12, no. 2, pp. 189-187, 2015.
- [8] A. F. Hidayatullah dan A. SN, "Analisis Sentimen dan Klasifikasi Kategori terhadap Tokoh Publik pada Twitter," dalam *Seminar Nasional Informatika (semnasIF)*, Yogyakarta, 2014.
- [9] Twitter, "Tweet object," [Online]. Available: <https://dev.twitter.com/overview/api/tweets>.
- [10] R. V. Kozinets, K. de Valck, A. C. Wojnicki dan S. J. Wilner, "Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities," *Journal of Marketing*, vol. 74, no. 2, pp. 71-89, 2010.
- [11] K. C. Laudon dan C. G. Traver, E-commerce, New Jersey: Pearson Education, Inc., 2014.
- [12] D. R. Kosiur, Understanding Electronic Commerce, Microsoft Press, 1997.
- [13] V. Vapnik dan C. Cortes, "Support Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.
- [14] C. W. Hsu, C. C. Chang dan C. J. Lin, "A Practical Guide to Support Vector Classification," Mei 2016. [Online]. Available: <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>. [Diakses Januari 2018].
- [15] A. Statnikov, D. Hardin, I. Guyon dan C. Aliferis, "A Gentle Introduction to Support Vector Machines in Biomedicine," 2009. [Online]. Available: <https://www.eecis.udel.edu/~shatkay/Course/papers/UOSVMAliferisWithoutTears.pdf>. [Diakses Januari 2018].
- [16] C. C. Chang dan C. J. Lin, "LIBSVM -- A Library for Support Vector Machines," Desember 2016. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.