

Komparasi Algoritma Klasifikasi Text Mining Untuk Analisa Sentimen Pada Akun Twitter Tokopediacare

Romy Triadi Sulisty, Raden Mirza Kurnia Fatoni, Dede Miftah Sururi, Emrina Telaumbanua, Gigih Hasoko Wibowo, Dedi Dwi Saputra
Universitas Nusa Mandiri
11213006@nusamandiri.ac.id

Abstrak

Perkembangan teknologi yang begitu pesat tidak terelakkan lagi. Media sosial kini menjadi sarana media komunikasi yang paling mudah digunakan orang banyak. Dengan media sosial, banyak penilaian sentimen yang dapat diteliti. Salah satunya Twitter. Dengan twitter, pengguna dapat melakukan review sebuah produk berdasarkan pengalaman yang mereka alami. Tokopedia misalnya, platform e-commerce dengan jutaan akun pembeli tentunya seringkali mendapatkan respon positif atau negatif. Melalui twitter, tokopedia menyediakan layanan konsumen dengan akun @TokopediaCare. Analisa sentimen terhadap review pengguna tokopedia pada akun tersebut memberikan indikator yang berguna untuk berbagai tujuan yang dapat ditemukan dalam komentar, umpan balik ataupun kritik. Data kicauan yang telah dikumpulkan dari twitter diolah terlebih dahulu dipecah menjadi kata sehingga dapat memudahkan dalam mengklasifikasi jenis kata. Penelitian ini menggunakan algoritma Synthetic Minority Oversampling Technique (SMOTE), Naive Bayes, Adaptive Boosting (AdaBoost), Support Vector Machine (SVM) dan Support Vector Machine Particle Swarm Optimization (SVM PSO) yang pengujiannya akan membandingkan dari campuran metode tersebut untuk mengetahui metode mana yang paling ideal dalam menentukan sentimen analisa pada kicauan twitter tokopedia. Dengan pengolahan hasil uji menggabungkan Synthetic Minority Oversampling Technique (SMOTE) dan Support Vector Machine Particle Swarm Optimization (SVM PSO) menghasilkan nilai terbaik yaitu Accuracy 76,05%, Precision 77,23%, Recall 74.14% dan AUC 0,826.

Kata kunci: Analisis Sentimen, Tokopedia, Naive Bayes, Support Vector Machine, Particle Swarm Optimization

Abstract

The rapid development of technology is unavoidable. Social media is now the easiest means of communication for many people to use. With social media, many sentiment assessments can be researched. One of them is Twitter. With Twitter, users can review a product based on their experiences. Tokopedia, for example, an e-commerce platform with millions of buyer accounts, of course, often gets a positive or negative response. Through Twitter, Tokopedia provides customer service with the @TokopediaCare account. Sentiment analysis of Tokopedia user reviews on these accounts provides useful indicators for various purposes which can be found in comments, feedback or criticism. The chirping data that has been collected from Twitter is first processed and broken down into words so that it can make it easier to classify the types of words. This study uses the Synthetic Minority Oversampling Technique (SMOTE) algorithm, Naive Bayes, Adaptive Boosting (AdaBoost), Support Vector Machine (SVM), and Support Vector Machine Particle Swarm Optimization (SVM PSO) whose tests will compare the mixture of these methods to find out which method the most ideal in determining sentiment analysis on Tokopedia's Twitter tweets. By processing the test results combining Synthetic Minority Oversampling Technique (SMOTE) and Support Vector Machine Particle Swarm Optimization (SVM PSO) the best values are Accuracy 76.05%, Precision 77.23%, Recall 74.14% and AUC 0.826.

Keywords : Sentiment Analysis, Tokopedia, Naive Bayes, Support Vector Machine, Particle Swarm Optimization

PENDAHULUAN

Indonesia menghadapi peningkatan pertumbuhan yang pesat di bidang e-commerce. Peningkatan ini tidak lepas dari besarnya jumlah pengguna internet di Indonesia (Gumilang, 2018). Selain itu, Indonesia adalah negara besar yang memiliki jumlah penduduk yang banyak sehingga membuat perkembangan e-commerce terus meningkat. Hasil survey pada tahun 2020 yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pengguna internet di Indonesia mencapai angka 196,7 juta orang yang sebelumnya pada tahun 2018 hanya 171,17 juta jiwa (Afdal et al., 2002).

Dari perkembangan tersebut pengguna internet dan media sosial merubah kebiasaan, salah satunya melakukan transaksi jual beli di marketplace seperti Tokopedia. Tokopedia merupakan toko online yang paling banyak dikunjungi masyarakat Indonesia yakni dengan total pengunjung mencapai 1,2 miliar yang terbagi dari 863,1 juta pengunjung dari web mobile dan 329,8 juta pengunjung dari desktop (Apriani & Gustian, 2019). Tokopedia menyediakan layanan pelanggan yang digunakan untuk menyampaikan pendapat, keluhan, dan pertanyaan dalam bertransaksi di Tokopedia melalui media sosial Twitter dengan akun @TokopediaCare. Twitter merupakan salah satu media sosial yang sangat populer di kalangan pengguna internet, hal ini dikarenakan kesederhanaan dan kemudahan dalam penggunaannya, serta pengguna dapat dengan bebas mengeluarkan pendapat atau opini mereka (Giovani et al., 2020). Beberapa Hal tersebut dapat dijadikan bahan analisis sentimen terhadap akun twitter @TokopediaCare.

Analisis sentimen adalah penggalan opini atau teks untuk mengklasifikasikan opini atau ulasan pengguna, dari sebuah ulasan merek, ulasan produk, atau ulasan layanan ke dalam kategori opini positif atau negative (Putri et al., 2020). Metode yang digunakan dalam penelitian ini adalah *Synthetic Minority Oversampling Technique (SMOTE)*, *Naïve Bayes*, *Adaptive Boosting (AdaBoost)*, *Support Vector Machine (SVM)* dan *Support Vector Machine Particle Swarm Optimization (SVM PSO)*.

Terdapat penelitian yang sudah dilakukan sebelumnya dengan mengklasifikasikan analisis

sentimen terhadap akun twitter seperti Irwansyah Saputra dan kawan-kawan dalam judul "Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning" membandingkan algoritma *Decision Tree*, *K-NN*, dan *Naïve Bayes* menghasilkan *Decision Tree* sebagai hasil tertinggi dengan akurasi 82%, presisi 81,95% dan recall 86% (Saputra et al., 2021).

Penelitian berikutnya ditulis Dinda dengan judul Komparasi Algoritma Klasifikasi *Text Mining* untuk Analisis Sentimen pada Review Restoran yang membandingkan algoritma *Naïve Bayes* dan *Support Vector Machine* dengan hasil algoritma *Naïve Bayes* lebih unggul dari *Support Vector Machine* dengan tingkat akurasi sebesar 87%. Sedangkan algoritma *Support Vector Machine* hanya menghasilkan akurasi 56% (Muthia, 2022). Dengan adanya perbedaan beberapa penelitian sebelumnya, mendorong penulis untuk melakukan penelitian dengan membandingkan beberapa metode klasifikasi dengan kasus yang berbeda. Di dalam penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan komparasi algoritma klasifikasi text mining untuk analisis sentimen terhadap Tokopedia pada media sosial Twitter.

Maksud dari penelitian ini adalah untuk mengetahui metode mana yang mendapatkan nilai *Accuracy*, *Precision*, *Recall* dan *Area Under the Curve (AUC)* yang terbaik antara algoritma *Naïve Bayes*, *Adaptive Boosting (AdaBoost)*, *Support Vector Machine (SVM)* dan *Support Vector Machine Particle Swarm Optimization (SVM PSO)* yang dioptimalkan dengan *Synthetic Minority Oversampling Technique (SMOTE)*.

METODE

Metode penelitian yang dilakukan adalah metode penelitian eksperimen, dengan tahapan sebagai berikut.

1. Pengumpulan Data
Pengumpulan data dengan menggunakan data yang bersumber dari komentar berbahasa Indonesia pada akun Twitter @TokopediaCare menggunakan RapidMiner Studio 9.10 untuk membuat dataset berisi 1.000 ulasan kemudian melakukan labeling

yang menghasilkan 428 ulasan positif dan 572 ulasan negatif.

Tabel 1. Contoh Dataset Twitter @TokopediaCare

No	Text	Label
1	@TokopediaCare Pake telur ikan mas hasil ayank blusukan ke pemancingan nyari ikan mas yang bunting. Makasih ayank @fadilahkiky	Positif
2	@TokopediaCare belanja pake voucher 90% trus ga dikonfirmasi seller, smp batal otomatis lalu vouchernya hangus. Mana dapetnya susahhh... tegaa	Negatif
3	wahai, si cepat. masa 6 hari barang masih nyangkut di bandung terus? tegal-bekasi gak jauh loh. dulu pesen dari jogja aja nyampe bisa 2-3 hari. gimana nih, maszeh/mbae? ?	Negatif

2. Pengolahan Data Awal

Pada pengolahan data awal dilakukan tahap preprocessing untuk mengubah data agar lebih mudah digunakan sebelum analisis sentimen (Andreyestha & Azizah, 2022) preprocessing dapat dilakukan dengan melalui beberapa proses yaitu :

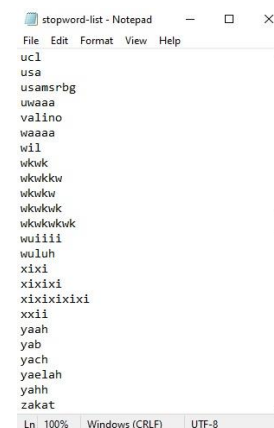
- a) *Annotation Removal* : Menghapus kata-kata imbuhan
- b) *Transformation - Remove URL* : Merubah dan menghapus URL seperti "http://" atau "https://".
- c) *Regex* : *Regular expression* atau yang sering disebut sebagai *Regex* adalah bahasa pemrograman kecil yang sangat khusus untuk pencarian pola

suatu kalimat atau string (Kuchling, 2022).

- d) *Tokenisasi* : menghilangkan tanda baca atau simbol yang bukan huruf.
- e) *Stemming* : kata-kata yang memiliki kata dasar yang sama akan dikelompokkan.
- f) *Stopword removal* : kata-kata yang tidak relevan seperti "wkwk", "yach", "yahh", dan sebagainya akan dihapus



Gambar 1. File Stemming

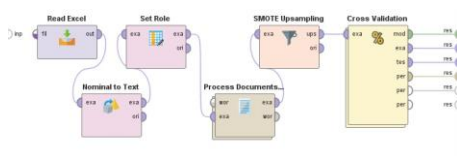


Gambar 2. File Stopword-list

3. Metode yang diusulkan

Untuk mencari hasil algoritma klasifikasi yang paling tinggi pada akun Twitter @TokopediaCare. Penulis menggunakan metode algoritma *Naïve Bayes*, *Adaptive Boosting (AdaBoost)*, *Support Vector Machine (SVM)* dan *Support Vector Machine Particle Swarm Optimization (SVM PSO)* yang dioptimalkan dengan *Synthetic Minority Oversampling Technique (SMOTE)*.

4. Eksperimen dan Pengujian Metode
Eksperimen pada model yang digunakan menggunakan Rapid Miner 9.10 untuk mengolah data. Model diuji untuk melihat hasil yang akan digunakan untuk mengambil keputusan dari hasil penelitian.

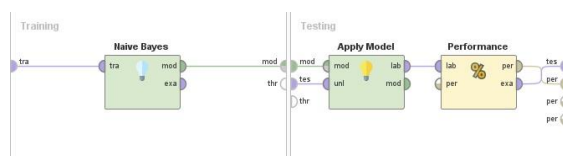


Gambar 3. *Design Proses Pengujian Menggunakan RapidMiner 9.10*

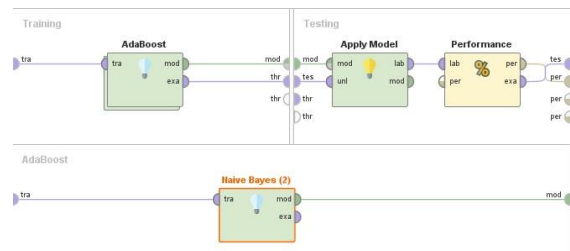
5. Evaluasi dan Validasi Hasil
Dalam sebuah penelitian dilakukan evaluasi terhadap model untuk mengetahui nilai *Accuracy*, *Precision*, *Recall* dan *AUC* dari model yang telah digunakan. Validasi hasil digunakan untuk melihat perbandingan dari model yang digunakan dengan hasil yang telah dilakukan sebelumnya.

HASIL DAN PEMBAHASAN

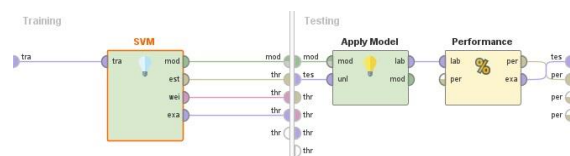
Penelitian ini didukung dengan aplikasi RapidMiner Studio 9.10 dengan menggunakan dataset kicauan dari akun twitter @TokopediaCare berjumlah 1.000 ulasan yang terdiri dari 428 ulasan positif dan 572 ulasan negatif. Pengujian dilakukan dengan menguji coba model usulan terhadap dataset. Model usulan yang diuji meliputi *Synthetic Minority Oversampling Technique (SMOTE)*, *Naïve Bayes*, *Adaptive Boosting (AdaBoost)*, *Support Vector Machine (SVM)* dan *Support Vector Machine Particle Swarm Optimization (SVM PSO)*. Pengujian terhadap algoritma tersebut dilakukan secara bergantian.



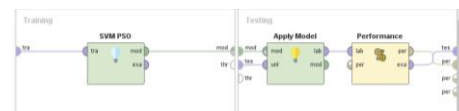
Gambar 4. *Menggunakan Algoritma Naïve Bayes Pada RapidMiner 9.10*



Gambar 5. *Menggunakan Algoritma Naïve Bayes dan Adaboost Pada RapidMiner 9.10*



Gambar 6. *Menggunakan Support Vector Machine (SVM) Pada RapidMiner 9.10*



Gambar 6. *Menggunakan Support Vector Machine Particle Swarm Optimization (SVM PSO) Pada RapidMiner 9.10*

Tabel 2. Hasil nilai komparasi algoritma

No	Algoritma	Accur acy	Preci sion	Recall	AUC
1	SMOTE + Naive Bayes	74.3	83.05	61.36	0.768
2	SMOTE + Naive Bayes + Adaboost	74.21	82.7	61.37	0.748
3	SMOTE + SVM	67.57	92.76	38.15	0.851
4	SMOTE + SVM PSO	76.05	77.23	74.14	0.826

KESIMPULAN DAN SARAN

Dari hasil pengujian beberapa metode yang berbeda yaitu *Synthetic Minority Oversampling Technique (SMOTE)*, *Naïve Bayes*, *Adaptive Boosting (AdaBoost)*, *Support Vector Machine (SVM)* dan *Support Vector Machine Particle*

Swarm Optimization (SVM PSO). Dengan pengolahan hasil uji menggabungkan *Synthetic Minority Oversampling Technique (SMOTE)* dan *Support Vector Machine Particle Swarm Optimization (SVM PSO)* menghasilkan nilai *Accuracy* 76,05%, *Precision* 77,23%, *Recall* 74.14% dan *AUC* 0,826. Sehingga penggunaan metode *Synthetic Minority Oversampling Technique (SMOTE)* dan *Support Vector Machine Particle Swarm Optimization (SVM PSO)* pada klasifikasi analisis sentimen akun twitter TokopediaCare terbukti memberikan hasil terbaik.

DAFTAR PUSTAKA

1. Gumilang, Z. A. N., dkk. (2018). Implementasi naive bayes classifier dan asosiasi untuk analisis sentimen data ulasan aplikasi e-commerce shopee pada situs google play.
2. Afdal M, Rahma Elita L, Studi Sistem Informasi P, Sains dan Teknologi UIN Suska Riau Jl Soebrantas KM FH, Pekanbaru -Riau P. PENERAPAN TEXT MINING PADA APLIKASI TOKOPEDIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *J Ilm Rekayasa dan Manaj Sist Inf* [Internet]. 2022 [cited 2022 Jun 22];8(1). Available from: <http://ejournal.uin-suska.ac.id/index.php/RMSI/article/view/16595>
3. Apriani R, Gustian D, Program S, Sistem I, Putra UN, Indonesia S, et al. ANALISIS SENTIMEN DENGAN NAÏVE BAYES TERHADAP KOMENTAR APLIKASI TOKOPEDIA. *J Rekayasa Teknol Nusa Putra* [Internet]. 2019 Sep 12 [cited 2022 Jun 22];6(1):54–62. Available from: <https://rekayasa.nusaputra.ac.id/article/view/86>
4. Giovani AP, Ardiansyah A, Haryanti T, Kurniawati L, Gata W. ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA
5. Putri DA, Kristiyanti DA, Indrayuni E, Nurhadi A, Hadinata DR. Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review. *J Phys Conf Ser* [Internet]. 2020 Nov 1 [cited 2022 Jun 22];1641(1):012085. Available from: <https://iopscience.iop.org/article/10.1088/1742-6596/1641/1/012085>
6. Saputra I, Aji Pambudi RS, Darono HE, Amsury F, Fahdia MR, Ramadhan B, et al. Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning. *Fakt Exacta* [Internet]. 2021 Feb 16 [cited 2022 Jun 22];13(4):200. Available from: https://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/view/7074/3723
7. Muthia DA. KOMPARASI ALGORITMA KLASIFIKASI TEXT MINING UNTUK ANALISIS SENTIMEN PADA REVIEW RESTORAN. *J Pilar Nusa Mandiri* [Internet]. 2018 Mar 15 [cited 2022 Jun 22];14(1):69–74. Available from: <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/92>
8. Andreyestha, Azizah QN. Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma SMOTE. *Infotek J Inform dan Teknol* [Internet]. 2022 Jan 30 [cited 2022 Jun 22];5(1):108–16. Available from: <http://ejournal.hamzanwadi.ac.id/index.php/infotek/article/view/4581>
9. Kuchling AM. Regular Expression HOWTO — Python 3.10.5 documentation [Internet]. [cited 2022 Jun 22]. Available from: <https://docs.python.org/3/howto/regex.html>