

Analisis Sentiment Instagram Menggunakan Metode Support Vector Machine (SVM) Berbasis Grid Search Algorithm (GSA)

Agus Salim^{*1}, Windu Gata², M Hilman Fakhri³, Cicih Sri Rhyu⁴, Arif Budiarto⁵

^{1,2,3,4,5}Program Studi Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta

E-mail: ^{*1}114002462@nusamandiri.ac.id, ²windu@nusamandiri.ac.id,

³hilman.mhz@nusamandiri.ac.id, ⁴14002466@nusamandiri.ac.id, ⁵

14002437@nusamandiri.ac.id

Abstrak

Instagram adalah sebuah aplikasi berbagi foto dan video yang memungkinkan pengguna mengambil foto, mengambil video, menerapkan filter digital, dan membagikannya ke berbagai layanan jejaring sosial, termasuk milik Instagram sendiri.

Pada penelitian ini akan dilakukan Analisis sentiment Instagram menggunakan metode support vector machine (SVM) berbasis Grid Search Algorithm (GSA). SVM salah satu metode yang dapat melakukan teknik klasifikasi kalimat menjadi positif, negatif ataupun netral, karena proses yang akan dilakukan bersifat non linear maka parameter yang akan digunakan adalah nilai C dan γ . Agar proses klasifikasi lebih optimal maka digunakan GSA sebagai model seleksi fitur. Untuk membuat sebuah aplikasi analisis sentimen diperlukan data training dan data testing. Dataset yang digunakan Sanders Instagram.

Dataset tersebut dilabel secara manual dan terdiri dari 654 negatif, 570 positif, 2503 netral, 1786 irrelevant. Tahap-tahap analisis sentimen dimulai dengan Loading data, Tokenizing, Weighting, Preprocessing, Filtering dan klasifikasi. Dari hasil uji coba, Analisis sentimen pada aplikasi memiliki tingkat keakuratan sekitar 79%. Persentase tag Instagram pada data sanders cenderung lebih banyak tag Instagram netral dan negatif dari pada positif.

Kata Kunci—Instagram, Analisis Sentiment Instagram Support Vector Machine

1. PENDAHULUAN

Instagram merupakan media sosial yang pada dasarnya berfungsi untuk berbagi foto maupun video pada sesama pengguna. Sejak September 2017 pengguna aktif instagram telah tercatat kurang lebih 800juta orang. Pengguna Instagram tentunya tak hanya orang dewasa namun anak-anak dan remaja ikut menggunakan media sosial ini.[2]

Analisis sentimen adalah proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks (dokumen, kalimat, sentimen, dll) bersifat positif, negatif atau netral. Analisis sentiment atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi yang terkandung dalam suatu kalimat opin[3]. Secara umum Ada dua pendekatan utama untuk melakukan analisis sentimen - machine learning base dan Lexicon base pembelajaran mesin didasarkan pada seleksi dan ekstraksi set sesuai fitur yang digunakan untuk mendeteksi sentimen. Sementara Lexicon base bergantung pada pra-definisikan kamus leksikon dan / atau dataset.

Sebelumnya telah dilakukan beberapa penelitian analisis sentimen terhadap Instagram dengan berbagai metode diantaranya:

1. Naïve Bayes (NB), Decission Tree (DT), k-Nearest Neighbor (KNN) pada bahasa Urdu. [5]

2. Support Vector Machine (SVM), NB, Maximun Entropy (ME) pada stock harga dan korelasinya. [1]
3. NB, SVM dan DT pada Reputasi Merk Mobile Phone Provider.[6]
4. NB terhadap Calon Presiden 2014. [4]

Dalam penelitian ini metode yang digunakan dalam analisis sentimen Instagram terhadap dataset sanders Instagram menggunakan SVM untuk mengklasifikasi kategori teks, tetapi masalah lain adalah sulitnya pemilihan fitur yang sesuai dan optimal pada bobot atribut yang digunakan sehingga menyebabkan tingkat akurasi menjadi rendah, hal ini dibutuhkan suatu metode untuk membantu pemilihan fitur agar hasil nilai sentiment lebih akurat. Berdasarkan uraian pada latar belakang dapat dirumuskan masalahnya adalah sebagai berikut:

1. Bagaimana cara kerja metode support vector machine berbasis Grid Search yang digunakan dalam analisis sentimen dari dataset sanders Instagram?
2. Bagaimana cara melakukan analisis sentimen kata kunci pada sanders pada Instagram secara otomatis ?

Berdasarkan rumusan masalah diatas, maka tujuan dari penelitian ini adalah membuat aplikasi analisis sentimen terhadap kata kunci pada dataset sanders pada Instagram secara otomatis dengan menggunakan metode klasifikasi SVM berbasis GSA.

Manfaat penelitian ini antara lain agar dalam pengklasifikasian data pada proses analisis sentimen Instagram dalam penentuan nilai akurasi lebih tinggi, sehingga keputusan dalam pemilihan kata produk yang populer akan lebih tepat..

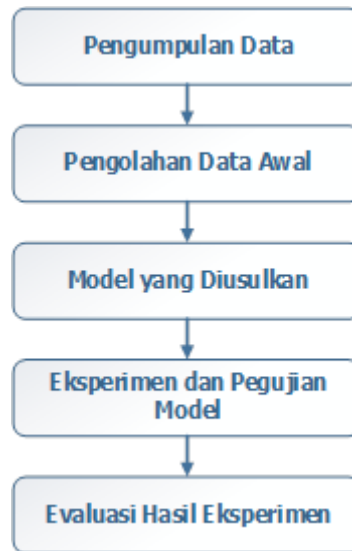
2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini untuk menentukan masalah, menentukan model dan metode, juga penjelasan bagaimana cara menganalisa sentiment Instagram dengan algoritma SVM berbasis GS sebagai pemilihan fitur terhadap kata kunci produk, pengumpulan data, pengklasifikasian, dengan algoritma SVM yang dilakukan seleksi atribut berbasis algoritma GS, diuraikan pula cara evaluasi dan validasi hasil penerapan, komparasi metode, metode analisis data, metode pengukuran penelitian, penerapan model dan metode pada masalah penelitian.

Dalam melakukan penelitian ini, penulis melakukan tahapan-tahapan kegiatan dengan mengikuti rencana kegiatan yang tertuang dalam kerangka berpikir sebagai berikut:

1. Pengumpulan data
Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data kedalam data set, termasuk variabel yang diperlukan dalam proses.
2. Pengolahan data awal
Ditahap ini dilakukan penyeleksian data, data dibersihkan dan ditransformasikan ke bentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model.
3. Metode yang diusulkan
Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (training data) dan data uji (testing data) juga diperlukan untuk pembuatan model.
4. Eksperimen dan pengujian metode
Pada tahap ini model yang diusulkan akan diuji untuk melihat hasil berupa rule yang akan dimanfaatkan dalam pengambilan keputusan.
5. Evaluasi dan validasi

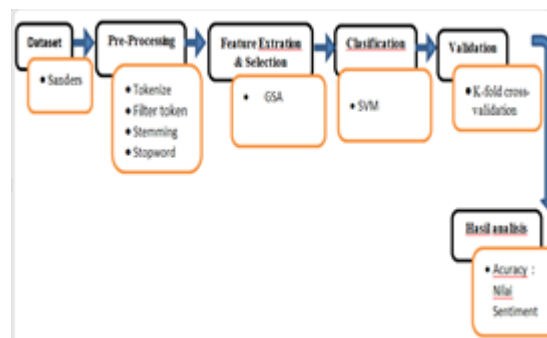
Pada tahap ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakurasian model.



Gambar 1. Alur Penelitian

2.1. Model yang Diusulkan

Pada tahap modeling ini dilakukan pemrosesan data traning sehingga akan membahas metode algoritma yang diuji dengan memasukkan data sentiment instagram diantaranya dataset STS-Test berupa kata kunci nama produk, perusahaan dan orang dan dataset sanders berupa kata kunci @apple, #google, #microsoft, # instagram, data kemudian di analisa dan dikomparasi. Berikut ini bentuk gambaran metode algoritma yang akan diuji.



Gambar 2. Model yang diusulkan

2.2. Eksperimen dan Pengujian Model

Tahap modeling untuk menyelesaikan analisis sentiment Instagram dengan menggunakan dua metode yaitu algoritma support vector machine sebagai metode klasifikasi data dan algoritma Grid search sebagai pemilihan fitur.

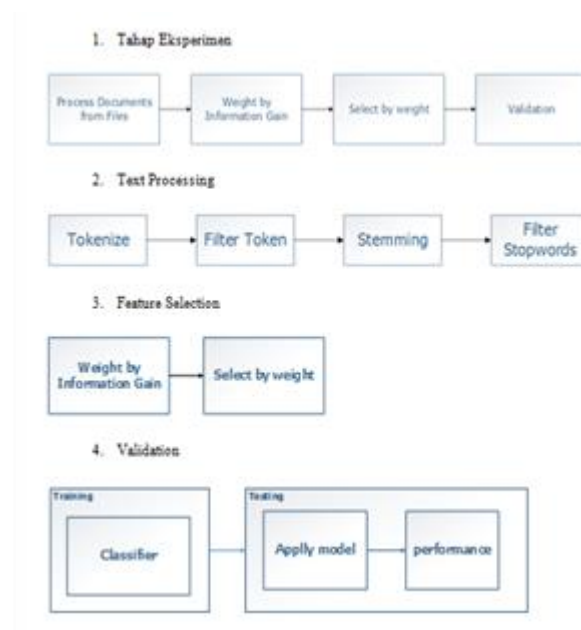
1. Support Vector Machine adalah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi (Maimon, 2010).
2. Grid Search Algorithm adalah salah satu prosedur pemilihan fitur model yang direkomendasikan pada metode classification (C.-W. Hsu, 2013).

Pada penelitian kali ini yang digunakan adalah penelitian Experiment. Penelitian eksperimen melibatkan penyelidikan hubungan kausal menggunakan tes dikendalikan oleh si peneliti itu sendiri. Dalam penelitian eksperimen digunakan spesifikasi software dan hardware sebagai alat bantu dalam penelitian pada Tabel 1:

Tabel 1 Spesifikasi Hardware dan Software

HARDWARE	RAM	4 GB
	CPU	Intel Core i3 460M CPU @2.53GHz
SOFTWARE	OS	MS Windows 7 Professional 64-bit
	APPS	Python canopy

Untuk memilih arsitektur pemodelan SVM dan GSA yang tepat, agar menghasilkan nilai akurasi dan nilai AUC yang terbesar, diperlukan pengaturan (adjustment) untuk tahapan pemodelan tersebut. Berikut ini adalah tahapan-tahapan yang membutuhkan adjustment:



Gambar 3. Tahapan Pemodelan.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pengujian Metode klasifikasi SVM Berbasis GA

Pada hasil pengujian nilai sentiment Instagram berdasarkan dataset Sanders diantaranya menggunakan SVM yang dikomparasi dengan metode MaxEntropy dan Naïve Bayes adalah sebagai berikut:

Accuracy 0.787587

Confusion Matrix

	n	e	p
	g	n	o
	a	e	s
	t	u	i
	i	t	t
	r	i	i
	v	a	v
	e	l	e
negative	<.>	359	.
neutral	.	<2373>	.
positive	.	281	<.>

(row = reference; col = test)

Gambar 4. Hasil Pengujian

3.2. Implikasi Penelitian

Hasil dari penelitian ini mempunyai implikasi terhadap penilaian produk yaitu meliputi implikasi terhadap aspek sistem pendukung keputusan penilaian produk yang paling populer, terhadap aspek manajerial dan terhadap aspek penelitian-penelitian selanjutnya yang akan diuraikan dibawah ini.

1. Implikasi terhadap aspek sistem pendukung keputusan penilaian produk. Hasil evaluasi menunjukkan penerapan Grid Search Algorithm untuk seleksi atribut dalam Support Vector Machine (SVM) dan dilakukan penyesuaian pada parameter C, e, population mampu menunjukkan aspek berupa atribut yang berpengaruh terhadap hasil dari penilaian produk. Dari hasil ini menunjukkan bahwa metode Support Vector Machine berbasis Grid Search merupakan metode yang cukup baik dalam pengklasifikasian data. Dengan demikian metode Support Vector Machine berbasis Grid Search dapat memberikan pemecahan permasalahan analisa sentiment terhadap twitter berupa produk yang paling populer dan dapat mendukung pengambilan keputusan dan pengembangan sistem informasi dengan menggunakan bantuan software Python Canopy
2. Implikasi terhadap aspek penelitian lanjutan. Dalam penelitian ini menggunakan metode Support Vector Machine (SVM) berbasis Grid Search, untuk mengetahui kehandalan metode, maka pada penelitian selanjutnya dapat dilakukan penggunaan data set lebih dari satu. Penelitian semacam ini dapat dikembangkan pada analisis serupa atau yang lain. Pengembangan dapat dilakukan dengan pemilihan parameter menggunakan metode Genetic Algorithm untuk meningkatkan hasil optimasi. Penggunaan algoritma yang lain juga dapat digunakan misalkan dengan metode Neural Network, C4.5, K- Nearest Neighbor dan sebagainya. Karena keputusan berdasarkan data sentiment twitter semakin berkembang maka kajian semacam ini dapat dilakukan secara periodik.

3.3. Pembahasan

Tingkat akurasi dalam penilaian sentiment terhadap Instagram merupakan masalah serius. Untuk menghasilkan nilai akurasi yang optimal maka diperlukan sebuah metode yang paling tepat. Pada penelitian ini Support Vector Machine (SVM) digunakan karena diketahui dari

hasil penelitian sebelumnya bahwa Support Vector Machine (SVM) memiliki kemampuan generalisasi yang sangat baik untuk memecahkan masalah walaupun dengan sampel yang terbatas. Eksperiment menggunakan metode support vector machine menghasilkan tingkat akurasi sebesar 70.30 %. Dari hasil tersebut diketahui bahwa keberhasilan dari Support Vector Machine (SVM) sangat dipengaruhi oleh pemilihan atribut yang tepat. Semakin banyak atribut dan informasi yang digunakan akan mengakibatkan banyaknya waktu dan biaya yang dikorbankan bahkan akan mengurangi tingkat akurasi dan kompleksitas yang lebih tinggi.

Mengingat pentingnya seleksi atribut dalam Support Vector Machine (SVM) maka diterapkan Grid Search Algorithm (GSA) untuk melakukan tugas tersebut. Grid Search Algorithm (GSA) diketahui dapat digunakan sebagai teknik optimasi untuk mengoptimalkan subset fitur. Algoritma GSA sederhana dan memiliki kompleksitas yang lebih rendah, sehingga dapat memastikan solusi optimal dengan menyesuaikan pencarian global dan lokal, sehingga kinerja klasifikasi Support Vector Machine (SVM) dapat ditingkatkan.

Eksperiment dilakukan kembali dengan menerapkan Grid Search Algorithm (GSA) untuk seleksi atribut dalam Support Vector Machine (SVM) dan dilakukan penyesuaian pada parameter C, dan population. Hasil dari eksperiment tersebut menunjukkan pengujian data STS-Test menggunakan Support Vector Machine (SVM) dan penerapan Grid Search Algorithm (GSA) dalam pemilihan atribut menunjukkan hasil yang lebih akurat dalam penentuan penilaian kredit ditandai dengan peningkatan nilai akurasi.

4. KESIMPULAN

Dengan menggunakan 5513 data training terdiri dari 654 negatif, 570 positif, 2503 netral, 1786 irrelevant, untuk mengklasifikasi data testing yang terdiri dari kfold cross = 10 maka SVM berbasis GSA memberikan hasil dengan akurasi hingga 79,00%.

Dengan menggunakan penambahan metode preprocessing akurasi dapat mencapai hingga 69,22%. Dengan metode stopwordremoval akurasi dapat mencapai hingga 73,74%. Dengan metode Grid Search dan Two Step GridSearch akurasi dapat mencapai hingga 73,89%. Sedangkan keseluruhan optimasi tersebut dapat memberikan hasil dengan akurasi hingga 80,33%..

DAFTAR PUSTAKA

- [1] Fatmawati, M. A. (2017). Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus : Akun Facebook Group iRaise Helpdesk). Jurnal CoreIT, 3(1), 24–30.
- [2] Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya, 2(11), 4704–4713.
- [3] Nirawana, I. W. S., & Indrawan, I. G. (2016). Analisis Sentimen Pada Review Film Dengan Menggunakan Algoritma Klasifikasi Naive Bayes Berdasarkan Term Objects Keywords. Seminar Nasional Pendidikan Teknik Informatika (SENAPATI 2016), (Senapati), 172–174.
- [4] Nurhuda, F., Widya Sihwi, S., & Doewes, A. (2016). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. Jurnal Teknologi & Informasi ITS smart, 2(2), 35. <https://doi.org/10.20961/its.v2i2.630>

- [5] Wahyuningsih, S., & Utari, D. R. (2018). Perbandingan Metode K-Nearest Neighbor , Naive Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit. Konferensi Nasional Sistem Informasi 2018, 619–623. Retrieved from <http://jurnal.atmaluhur.ac.id/index.php/knsi2018/article/view/424/349>
- [6] Windarti, M., & Suradi, A. (2019). Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Studi Mahasiswa. Telematika, 12(1), 14. <https://doi.org/10.35671/telematika.v12i1.778>