

Analisis Sentimen dengan Menggunakan Metode Klasifikasi Lazy K-Star

Nurirwan Saputra

Program Studi Teknik Informatika, Universitas PGRI Yogyakarta

Email: nurirwan@upy.ac.id

Abstrak

Analisis Sentimen merupakan cabang ilmu dari text mining yang sudah mulai banyak dimanfaatkan dalam berbagai bidang, misalnya dalam bidang politik, bidang pemasaran produk, bidang jasa, bidang perhotelan dan lain lain. Contoh analisis sentimen dalam dunia bisnis yaitu digunakan untuk menganalisis secara otomatis opini pelanggan tentang produk dan pelayanannya. Dalam bidang politik, analisis sentimen digunakan untuk menampung aspirasi masyarakat mengenai kinerja pemerintahan maupun manajemen reputasi tokoh tertentu. Dalam bidang pendidikan, termasuk dalam penelitian ini, dengan melakukan analisis sentimen mahasiswa terhadap sebuah Universitas atau Perguruan Tinggi, dapat menjadi evaluasi Perguruan Tinggi tersebut dalam meningkatkan kualitas pembelajarannya.

Penelitian ini menggunakan data primary yang berasal dari mahasiswa yang minimal sudah melakukan pembelajaran minimal setahun di Universitas PGRI Yogyakarta, data yang didapat, akan dilakukan normalisasi kata terlebih dahulu, sehingga kata yang tidak baku menjadi baku. Kemudian penelitian ini menggunakan Word Tokenizer dengan menggunakan metode klasifikasi Lazy K-Star.

Hasil klasifikasi menggunakan metode Lazy K-Star tidak sebaik dengan menggunakan metode Naive Bayes dan SVM dengan persentase sebesar 79,2952 %, sehingga perlu adanya improvisasi atau pembelajaran tambahan yang dilakukan guna mendapatkan hasil akurasi yang lebih baik.

Kata-kata kunci: analisis sentimen, Lazy K-Star

Pendahuluan

Sentiment analysis atau opinion mining merupakan topik riset yang penting dan sedang marak dilakukan saat ini. Opinion mining merupakan cabang penelitian dari text mining. Fokus dari opinion mining adalah melakukan analisis opini dari suatu dokumen teks [1]. Saat ini, pendapat khalayak umum menjadi sumber yang penting dalam pengambilan keputusan [2]. Secara umum terdapat dua tipe informasi tekstual di web yaitu fakta dan opini. Fakta adalah pernyataan objektif mengenai entitas dan kejadian di dunia sedangkan opini adalah pernyataan subjektif yang merefleksikan sentimen atau persepsi orang mengenai entitas ataupun kejadian di dunia. Ketika suatu organisasi/perusahaan/perorangan ingin memperoleh opini publik mengenai produk, citra dan layanannya maka mereka tidak perlu melakukan survei konvensional dan fokus group yang mahal biayanya [3]. Begitu pula dalam

organisasi pendidikan khususnya universitas, dengan menggunakan analisis sentimen, dapat menjadi evaluasi dalam meningkatkan kualitas pendidikan dalam universitas, sehingga lulusan yang dihasilkan menjadi lebih berkualitas.

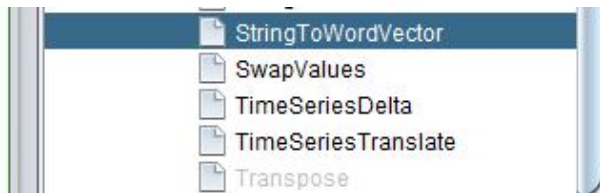
Pada penelitian ini, Mahasiswa merupakan aset utama di sebuah Universitas, baik buruknya sebuah Universitas, tergantung dari mahasiswanya, semakin baik lulusan sebuah Universitas, maka semakin baik pula citra sebuah Universitas, dan semakin banyak pula mahasiswa baru yang berminat masuk di Universitas tersebut. Selain itu juga, proses belajar dan mengajar yang terjadi selama masa perkuliahan sangat berpengaruh terhadap kenyamanan dan hasil yang diperoleh mahasiswa. Maka dari itu, mahasiswa merupakan aset penting dalam mengembangkan sebuah Universitas. Tanpa adanya perhatian Universitas terhadap mahasiswanya, maka sebuah Universitas akan sulit berkembang [4].

Teori

a. Unsupervised Learning

Dalam penelitian ini, menggunakan Unsupervised learning, yaitu melakukan pembelajaran tanpa bantuan guru. Dengan menggunakan unsupervised learning, ada ukuran yang jelas tentang keberhasilan atau kekurangan yang dapat digunakan untuk menilai situasi tertentu dan untuk membandingkan keefektifan sebuah metode dalam berbagai situasi [5].

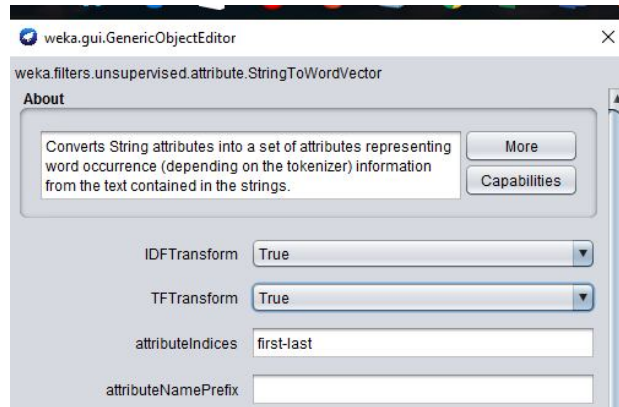
StringToWordVector diperlukan untuk mengkonversi string atribut dalam satu set atribut yang mewakili kemunculan suatu kata [6]. Pada Gambar 1 memperlihatkan Unsupervised learning dengan mengubah String menjadi Word Vector (StringToWordVector).



Gambar 1. Menggunakan Unsupervised Learning dengan mengubah String menjadi Word Vector

b. TF-IDF

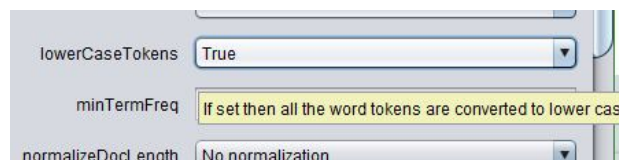
TF-IDF menghitung nilai-nilai untuk setiap kata dalam sebuah dokumen berbanding terbalik dari frekuensi kemunculan kata dalam sebuah dokumen tertentu dengan persentase kemunculan sebuah dokumen. Kata-kata dengan angka TF-IDF yang tinggi menyiratkan hubungan yang kuat dengan dokumen dimana kata tersebut muncul [7]. TF-IDF secara default pada weka adalah false, sehingga perlu diubah menjadi true seperti pada Gambar 2.



Gambar 2. TF-IDF True

c. Lower Case

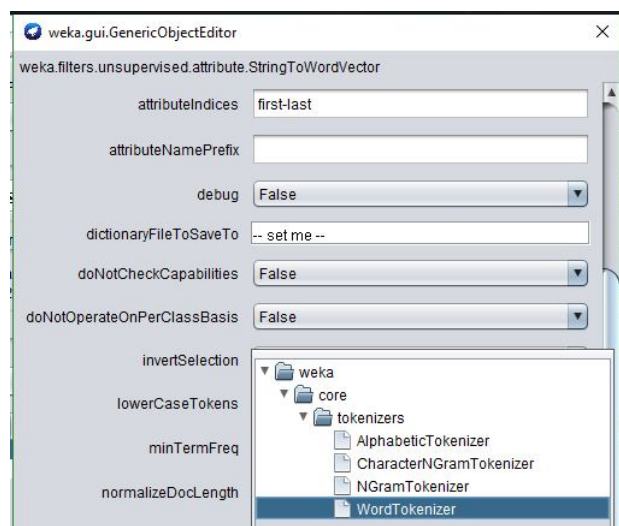
Dengan mengubah tiap huruf menjadi huruf kecil atau lower case pada Gambar 3, tiap kata yang masuk dengan ukuran font yang berbeda dapat dikenali sebagai kata yang sama maknanya.



Gambar 3. Mengubah semua huruf menjadi lower case

d. Word Tokenizer

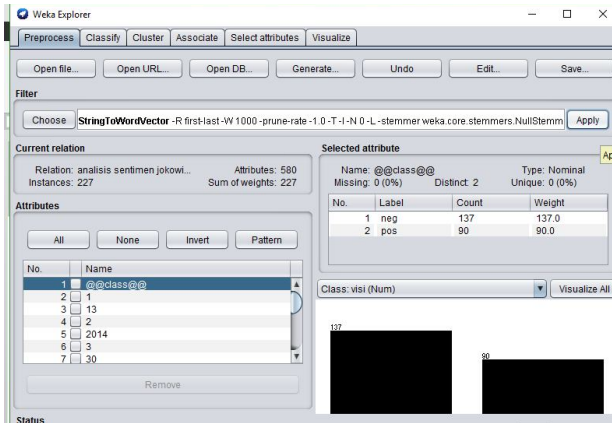
Dengan menggunakan Word Tokenizer seperti pada Gambar 4, dapat membagi kalimat berdasarkan pembatas yaitu spasi [8]. Sehingga setiap kata akan diakhiri spasi, kemudian dilanjutkan kata yang baru.



Gambar 4. Menggunakan Word Tokenizer

e. Apply

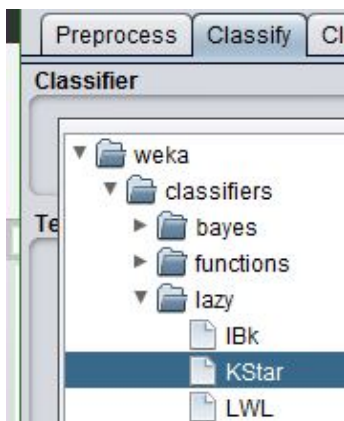
Dengan menekan tombol “Apply” pada Gambar 5, semua pengaturan akan diimplementasikan, dan akan terlihat hasil dari Word Tokenizer serta jumlah data yang dimasukkan.



Gambar 5. Apply pengaturan, dan terlihat data negatif sebanyak 137 dan positif sebanyak 90

f. Metode Klasifikasi Lazy KStar

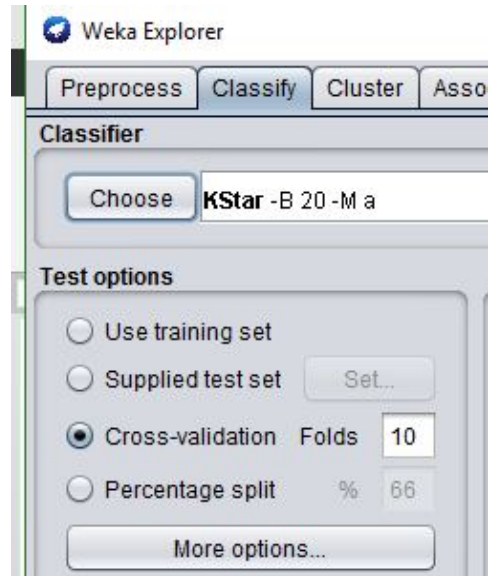
Metode klasifikasi Lazy K-Star menyimpan data training dan tidak melakukan operasi sesungguhnya hingga waktu klasifikasi berlangsung, Lazy K-Star mendukung pembelajaran tambahan [9]. Pada WEKA, pemilihan Metode Klasifikasi Lazy K-Star dapat dilakukan dengan masuk pada tab Classify kemudian pada Classifier memilih Lazy kemudian KStar seperti pada Gambar 6.



Gambar 6. Klasifikasi menggunakan Metode Lazy KStar

g. 10 Fold Cross Validation

Proses validasi menggunakan 10-fold cross validation pada Gambar 7, adalah dimana data dibagi secara acak menjadi 10 bagian data dengan jumlah yang sama. Sehingga dilakukan proses validasi sebanyak 10 kali secara berulang. Sedangkan untuk tingkat kebenaran proses klasifikasi ditabulasikan dalam suatu Tabel yang disebut confusion matrix [10]



Gambar 7. Menggunakan test option 10 Fold Cross-Validation

Hasil dan diskusi

Hasil yang didapat pada penelitian analisis sentimen menggunakan metode klasifikasi Lazy K-Star tidak sebaik dengan menggunakan metode klasifikasi Naïve Bayes dan SVM dengan akurasi sebesar 79,2952 % seperti yang ditunjukkan pada Gambar 8.

Select attributes	Visualize
Classifier output	
Time taken to build model: 0 seconds	
=== Stratified cross-validation ===	
=== Summary ===	
Correctly Classified Instances	180 79.2952 %
Incorrectly Classified Instances	47 20.7048 %
Kappa statistic	0.5855
Mean absolute error	0.2449
Root mean squared error	0.3979
Relative absolute error	51.1571 %
Root relative squared error	81.343 %
Total Number of Instances	227
=== Detailed Accuracy By Class ===	
	TP Rate FP Rate Precision Recall F-Measure MCC
	0,745 0,133 0,895 0,745 0,813 0,598
	0,867 0,255 0,690 0,867 0,768 0,598
Weighted Avg.	0,793 0,182 0,814 0,793 0,795 0,598
=== Confusion Matrix ===	
a b <-- classified as	
102 35 a = neg	
12 78 b = pos	

Gambar 8. Hasil Klasifikasi

Kesimpulan

Tuliskan hasil kesimpulan Anda dalam bagian ini. Singkat saja tetapi jelas. Jangan mengulang terlalu banyak hal-hal pada bagian Hasil dan diskusi, akan tetapi rangkumkan. Bagian ini cukup satu paragraf saja.

Ucapan terima kasih

Penulis mengucapkan terima kasih atas Organisasi A atas dukungan finansialnya pada penelitian ini dan Organisasi B atau dukungannya dalam keikutsertaan dalam kegiatan ilmiah ini. Penulis juga berterima kasih kepada Ahli C atas dikusinya yang bermanfaat.

Referensi

- [1] I. F. Rozi, S. H. Pramono, and E. A. Dahlan, "Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *J. EECCIS*, vol. 6, no. 1, pp. 37–43, Mar. 2013.
- [2] V. Chandani, R. S. Wahono, and Purwanto, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature

Selection pada Analisis Sentimen Review Film," *J. Intell. Syst.*, vol. 1, no. 1, pp. 56–60, Feb. 2015.

- [3] G. A. Buntoro, T. B. Adji, and A. E. Permanasari, "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation." [Online]. Available: https://www.researchgate.net/profile/Ghulam_Buntoro/publication/270794878_Sentiment_Analysis_Twitter_dengan_Kombinasi_Lexicon_Based_dan_Double_Propagation/links/54b4f8e10cf2318f0f97037e.pdf. [Accessed: 13-Apr-2017].
- [4] N. Saputra, T. B. Adji, and A. E. Permanasari, "ANALISIS SENTIMEN DATA PRESIDEN JOKOWI DENGAN PREPROCESSING NORMALISASI DAN STEMMING MENGGUNAKAN METODE NAIVE BAYES DAN SVM," *J. Din. Inform.*, vol. 5, no. 1, Jun. 2015.
- [5] T. Hastie, R. Tibshirani, and J. Friedman, "Unsupervised Learning," in *The Elements of Statistical Learning*, Springer New York, 2009, pp. 485–585.
- [6] P. Han, D. B. Wang, and Q. G. Zhao, "The research on Chinese document clustering based on WEKA," in *2011 International Conference on Machine Learning and Cybernetics*, 2011, vol. 4, pp. 1953–1957.
- [7] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," *Dep. Comput. Sci. Rutgers Univ.*, 2003.
- [8] N. Mishra and A. Mishra, "Part of Speech Tagging for Hindi Corpus," in *2011 International Conference on Communication Systems and Network Technologies*, 2011, pp. 554–558.
- [9] S. D. Thepade and M. M. Kalbhor, "Novel data mining based image classification with Bayes, Tree, Rule, Lazy and Function Classifiers using fractional row mean of Cosine, Sine and Walsh column transformed images," in *2015 International Conference on Communication, Information*

Computing Technology (ICCICT), 2015, pp. 1–6.

- [10] N. D. Putranti and E. Winarko, “Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine,” *IJCCS Indones. J. Comput. Cybern. Syst.*, vol. 8, no. 1, pp. 91–100, Jan. 2014.

Nurirwan Saputra

Program Studi Teknik Informatika, Fakultas Teknik

Universitas PGRI Yogyakarta

nurirwan@upy.ac.id