

Sentiment Analysis pada Data Twitter dengan Pendekatan Naïve Bayes Multinomial

Bagus Setya Rintyarna¹⁾

¹⁾Jurusan Teknik Elektro, Fakultas Teknik, Universitas Muhammadiyah Jember
Jl. Karimata No. 49 Jember Kode Pos 68121
Email : ¹⁾bagus.setya@unmuhjember.ac.id

ABSTRAK

Sebagai platform di mana user bertukar informasi dalam bentuk pesan pendek, link ke website lain, gambar maupun video, twitter telah berevolusi menjadi platform microblogging yang menjadi sumber informasi bagi banyak permasalahan karena karakteristiknya yang bersifat real time. Salah satu informasi yang penting diekstraksi dari data twitter adalah opini atau sentimen. Teknik untuk mengekstraksi sentimen dengan pendekatan komputasional disebut sebagai *Sentiment Analysis*. Penelitian ini mengusulkan eksperimen dengan teknik *Sentiment Analysis* pada dataset twitter dengan metode Multinomial Naïve Bayes untuk kategorisasi data teks. Tool yang dipergunakan adalah WEKA. Pada penelitian ini dievaluasi pengaruh algoritma stemming yang berbeda, yaitu : 1) IteratedLovinsStemmer, 2) LovinsStemmer, 3) NullStemmer dan 4) SnowBallStemmer. Evaluasi kinerja disajikan dalam enam matriks parameter yaitu TP Rate, FP Rate, Precision, Recall dan F Measure serta ROC.

Kata Kunci : *Sentiment Analysis, Multinomial Naïve Bayes*

1. PENDAHULUAN

Microblog adalah platform di mana user bertukar informasi dalam bentuk pesan pendek, link ke website lain, gambar maupun video (Cámara, Valdivia, López, & Ráez, 2012). Platform *microblogging* telah berevolusi menjadi sumber informasi bagi banyak permasalahan karena karakteristiknya yang bersifat real time (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011). Twitter menjadi sarana bertukar opini pada beragam topik terkini dalam beragam bidang dan issue dalam kehidupan sehari-hari. Seperti mengekspresikan sentimen positif atas suatu produk atau sentimen negatif untuk seorang tokoh politik. Pendekatan untuk mengekstraksi informasi opini yang terkandung dalam data text dari platform twitter secara otomatis dengan pendekatan komputasional menjadi

penting dilakukan (Saif, He, Fernandez, & Alani, 2014).

Pendekatan untuk mengekstraksi informasi opini yang terkandung dalam suatu data text disebut sebagai *Sentiment Analysis* (Pang & Lee, 2008). *Sentiment analysis* termasuk bidang penelitian di bawah Natural Language Processing. Karena berkaitan dengan natural language, upaya untuk meningkatkan kinerja sentiment analysis bisa dengan mengekstraksi aspek kontekstual dalam data teks. Pendekatan yang dilakukan bisa dengan teknik *word sense disambiguation* (Rintyarna & Sarno, 2016). Untuk pendekatan menggunakan lexicon berbahasa Inggris seperti SentiwordNet, diperlukan metode machine translation, salah satunya dengan teknik parallel corpus (Suryani et al., 2016). Teknik yang dipergunakan untuk melakukan *Sentiment Analysis* dapat dikelompokkan menjadi dua jenis :

yaitu teknik berbasis Machine Learning seperti yang digunakan dalam penelitian ini dan teknik berbasis lexicon (Medhat, Hassan, & Korashy, 2014).

Sebagian besar penelitian *sentiment analysis* fokus pada dataset product review sehingga fokus penelitian berkaitan dengan *sentiment analysis* pada data twitter menjadi penting dilakukan dalam masa mendatang (Kumar & Sebastian, 2012). Penelitian ini mengusulkan eksperimen dengan teknik *Sentiment Analysis* pada dataset twitter. Mengingat besarnya jumlah dataset yang dipergunakan untuk eksperimen, penyediaan *ground truth* tidak dilakukan anotasi secara manual. Pendekatan yang dilakukan untuk eksperimen menggunakan metode *Multinomial Naïve Bayes for Text Categorization* dengan menggunakan WEKA. Untuk evaluasi kinerja, dilakukan 10 *folds cross validation* dengan parameter *precision*, *recall* dan *f-measure*.

2. TINJAUAN PUSTAKA

Sentiment Analysis pada dataset twitter pernah dilakukan oleh (Pak & Paroubek, 2010). Dalam penelitian ini diusulkan teknik untuk mengumpulkan dataset twitter dengan kelas positive dan negative tanpa perlu melakukan anotasi manual. Pendekatan yang dilakukan adalah dengan menghitung *word frequency*. Selanjutnya TreeTagger digunakan untuk mengetahui distribusi tag dalam dataset twitter. Dalam eksperimen, klasifikasi dilakukan dengan teknik machine learning yaitu SVM, CRF dan Naïve Bayes.

Kouloumpis (Kouloumpis, Wilson, & Moore, 2011) juga melakukan eksperimen pada dataset twitter. Dalam penelitian ini diinvestigasi peran fitur linguistik dalam teknik *sentiment analysis* menggunakan machine learning. Untuk *development* dan

training, digunakan HASH¹ dari Edinburgh Twitter Corpus dan EMOT². Sedangkan untuk evaluasi digunakan ISIEVE³. Fitur yang digunakan bervariasi, termasuk unigram dan bigram. Selain itu juga digunakan fitur sentimen dari sentiment lexicon dan fitur POS. Taboada (Taboada, Brooke, & Tofiloski, 2011) mengusulkan pendekatan berbasis lexicon yang disebut sebagai Semantic Orientation Calculator (SO-CAL). Orientasi semantic (SO) didefinisikan sebagai ukuran subyektifitas dan opini dalam suatu dokumen teks. SO menyatakan nilai positive dan negative serta potency atau derajatnya. Data yang digunakan dalam proses evaluasi adalah Amazon's Mechanical Turk Service⁴. Hasil eksperimen menunjukkan bahwa SO-CAL lebih baik dibanding beberapa metode baseline yang digunakan dalam evaluasi.

Dalam penelitian lain (Mullen & Collier, 2011) menggunakan teknik SVM untuk *sentiment analysis*. Penelitian ini mengevaluasi kinerja SVM dengan fitur unigram dikombinasikan dengan beberapa fitur lain yaitu : 1) orientasi semantic dengan PMI, 2) Osgood semantic differentiation, 3) topic proximity dan syntactic relation. SVM diimplementasikan dengan Kudo's Tiny SVM⁵.

3. METODE PENELITIAN

Teknik machine learning pada umumnya terdiri dari dua tahap yaitu tahap training dan tahap klasifikasi.

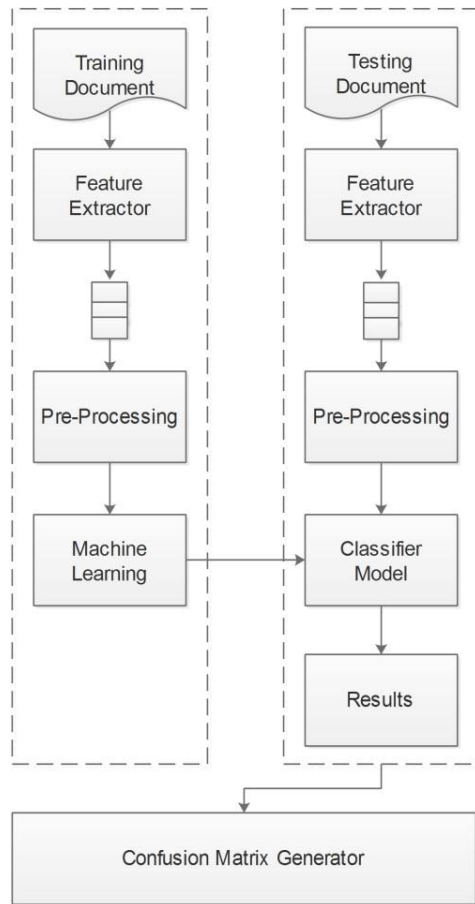
¹<http://demeter.inf.ed.ac.uk>

²<http://twittersentiment.appspot.com>

³<http://i-sieve.com>

⁴<http://mturk.com>

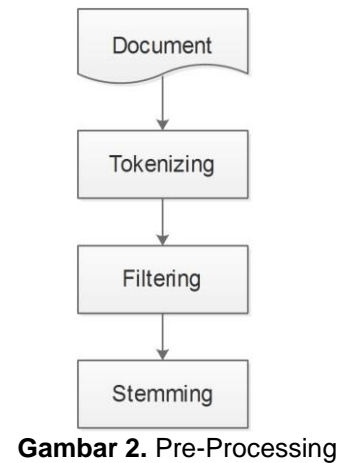
⁵<http://chasen.org/~taku/software/TinySVM/>



Gambar 1. Tahapan Klasifikasi Machine Learning

Tahap training disebut juga tahap *supervised learning* di mana machine learning algorithm yang digunakan berusaha mendapatkan model dari data yang dilabeli secara manual. Tahap berikutnya adalah tahap klasifikasi. Secara sederhana tahapan tersebut dapat digambarkan seperti Gambar 1.

Teknik *machine learning* yang digunakan dalam penelitian ini adalah Naïve Bayes Multinomial Text. Untuk proses klasifikasi, fitur yang di-generate adalah fitur *bag of word*. Untuk mendapatkan fitur *bag of word* dilakukan pre-processing seperti terlihat pada Gambar 2.



Gambar 2. Pre-Processing

Pada umumnya, dari *bag of word* dihitung TFIDF setiap kata. Formula untuk menghitung TFIDF dapat dilihat pada equation 1.

$$TFID(word) = \log(f + 1) \times \log\left(\frac{D}{df}\right) \quad (1)$$

Di mana f adalah frekuensi kata, D adalah jumlah total dokumen dan df adalah jumlah dokumen yang mengandung kata (*word*).

Sementara itu, Multinomial Naïve Bayes (MNB) menghitung probabilitas kelas untuk dokumen tertentu. Jika N adalah jumlah kata, C adalah set kelas dalam eksperimen maka MNB berusaha menghitung nilai probabilitas tertinggi untuk dokumen t_i dalam kelas C . MNB didefinisikan sebagaimana terlihat dalam equation 2.

$$\Pr(c|t_i) = \frac{\Pr(c)\Pr(t_i|c)}{\Pr(t_i)}, \quad c \in C \quad (2)$$

Paramater yang nantinya digunakan dalam penelitian ini untuk mengevaluasi kinerja algoritma yang digunakan adalah : TP Rate, FP Rate, Precision, Recall dan F Measure serta ROC. Confusion matrix yang menggambarkan hasil pengujian suatu data uji dapat dilihat pada Tabel 1. Dimana ada 4 kondisi hasil pengujian data

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

Tabel 1. Confusion Matrix untuk Evaluasi Performance

		Test Outcome	
		Positive	Negative
True Condition	Positive	TP	FP
	Negative	FN	TN

Dataset yang dipergunakan untuk mensimulasikan *sentiment analysis* pada data twitter dengan teknik Multinomial Naïve Bayes ini adalah dataset yang di-crawling dan dilabeli oleh Mark Hall⁶ menjadi kelas positive dan kelas negative di mana jumlah kelas positifnya sebanyak 41.403 data dan kelas negatifnya sebanyak 8.552 data. Proses pelabelan tidak dilakukan secara manual.

4. HASIL DAN PEMBAHASAN

Eksperimen pada penelitian ini dilakukan untuk mengevaluasi pengaruh penggunaan algoritma stemming yang berbeda yaitu: 1) Null Stemmer, 2) Iterated Lovins Stemmer, 3) Lovins Stemmer dan 4) Snow Ball Stemmer. Eksperimen dilakukan dengan 10-folds cross validation. Hasil eksperimen dengan Null Stemmer disajikan dalam Tabel 2 sampai Tabel 4.

Tabel 2. Detail Akurasi per Kelas dengan Null Stemmer

Class	TP-Rate	FP-Rate	Prec	Rec	F-Meas	ROC Area
Pos	0.857	0.424	0.907	0.857	0.882	0.815
Neg	0.576	0.143	0.455	0.576	0.508	0.815
Avg	0.809	0.376	0.830	0.809	0.818	0.815

Tabel 3. Confusion Matriks dengan Null Stemmer

		Test Outcome	
		Positive	Negative
True Condition	Positive	35499	5904
	Negative	3624	4928

Tabel 4. Summary Hasil dengan Null Stemmer

Parameter	Hasil
Kappa Statistic	0.3922
Mean Absolute Error	0.2441
Root Mean Squared Error	0.3738

Selanjutnya disajikan hasil eksperimen dengan Iterated Lovins Stemmer dalam Tabel 5 sampai Tabel 7.

Tabel 5. Detail Akurasi per Kelas dengan Iterated Lovins Stemmer

Class	TP-Rate	FP-Rate	Prec	Rec	F-Meas	ROC Area
Pos	0.899	0.504	0.896	0.899	0.898	0.816
Neg	0.496	0.101	0.503	0.496	0.499	0.816
Avg	0.830	0.435	0.829	0.830	0.829	0.816

Tabel 6. Confusion Matriks dengan Iterated Lovins Stemmer

		Test Outcome	
		Positive	Negative
True Condition	Positive	37220	4183
	Negative	4314	4238

Tabel 7. Summary Hasil dengan Iterated Lovins Stemmer

Parameter	Hasil
Kappa Statistic	0.3969
Mean Absolute Error	0.2257
Root Mean Squared Error	0.3526

Hasil eksperimen berikutnya menggunakan Lovins Stemmer disajikan dalam Tabel 8 sampai Tabel 10.

Tabel 8. Detail Akurasi per Kelas dengan Lovins Stemmer

Class	TP-Rate	FP-Rate	Prec	Rec	F-Meas	ROC Area
Pos	0.895	0.494	0.898	0.895	0.897	0.817
Neg	0.506	0.105	0.500	0.506	0.503	0.817
Avg	0.829	0.427	0.830	0.829	0.829	0.817

Tabel 9. Confusion Matriks dengan Lovins Stemmer

		Test Outcome	
		Positive	Negative
True Condition	Positive	37073	4330
	Negative	4222	4330

⁶<https://docs.google.com/open?id=0B1pvkpCwTsiSd1pyTFZkdWVRdEs5Q1NiQW1mRmF1Zw>

Tabel 10. Summary Hasil dengan Lovins Stemmer

Parameter	Hasil
Kappa Statistic	0.3997
Mean Absolute Error	0.2261
Root Mean Squared Error	0.3536

Dan terakhir eksperimen dilakukan dengan Snowball Stemmer yang hasilnya disajikan dalam Tabel 11 sampai Tabel 13.

Tabel 11. Detail Akurasi per Kelas dengan Snowball Stemmer

Class	TP-Rate	FP-Rate	Prec	Rec	F-Meas	ROC Area
Pos	0.857	0.424	0.907	0.857	0.882	0.815
Neg	0.576	0.143	0.455	0.576	0.508	0.396
Avg	0.809	0.376	0.830	0.809	0.818	0.815

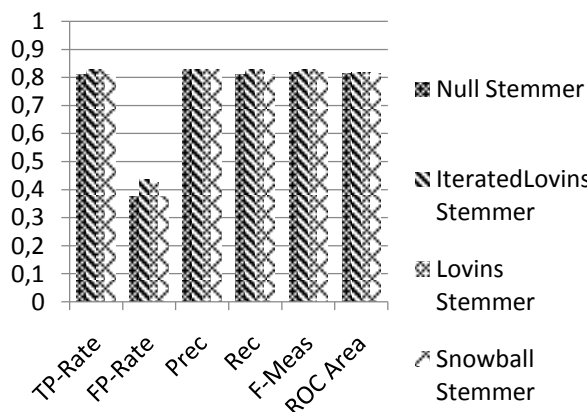
Tabel 12. Confusion Matriks dengan Snowball Stemmer

		Test Outcome	
		Positive	Negative
True Condition	Positive	35499	5904
	Negative	3624	4928

Tabel 13. Summary Hasil dengan Snowball Stemmer

Parameter	Hasil
Kappa Statistic	0.3922
Mean Absolute Error	0.2441
Root Mean Squared Error	0.3738

Perbandingan rata-rata kinerja masing-masing metode disajikan dalam Gambar 3.



Gambar 3. Perbandingan rata-rata kinerja

5. KESIMPULAN DAN SARAN

Dari hasil pengujian dapat disimpulkan beberapa hal sebagai berikut:

1. Naïve Bayes Multinomial adalah salah satu algoritma machine learning yang dapat digunakan untuk sentiment analysis dengan teknik klasifikasi data teks.
2. Meskipun berpengaruh dalam kinerja sentiment analysis, Null Stemmer, Iterated Lovins Stemmer, Lovins Stemmer dan Snowball Stemmer tidak menghasilkan kinerja yang berbeda secara signifikan.
3. Kinerja parameter paling rendah adalah FP-Rate.

DAFTAR PUSTAKA

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment Analysis of Twitter Data. In *Proc. ACL 2011 Workshop on Languages in Social Media*, (pp. 30–38).
- Cámara, E. M., Valdivia, M. T. M., López, L. A. U., & Ráez, A. R. M. (2012). Sentiment analysis in Twitter. *Natural Language Engineering*. <http://doi.org/10.1017/S1351324912000332>
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! *Proceedings of the Fifth International AAI Conference on Weblogs and Social Media (ICWSM 11)*, 538–541. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2857/3251?iframe=true&width=90%25&height=90%25>
- Kumar, A., & Sebastian, T. M. (2012). Sentiment Analysis on Twitter. *International Journal of Computer Science Issues*, 9(4), 372–378.

- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <http://doi.org/10.1016/j.asej.2014.04.011>
- Mullen, T., & Collier, N. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.
- Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *In Proceedings of the Seventh Conference on International Language Resources and Evaluation* (pp. 1320–1326). <http://doi.org/10.1371/journal.pone.0026624>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis, 2, 1–135. <http://doi.org/10.1561/1500000001>
- Rintyarna, B. S., & Sarno, R. (2016). Adapted Weighted Graph for Word Sense Disambiguation. In *2016 4th International Conference on Information and Communication Technology (ICoICT)* (Vol. 4, pp. 60–64). <http://doi.org/10.1109/ICoICT.2016.7571884>
- Saif, H., He, Y., Fernandez, M., & Alani, H. (2014). Contextual semantics for sentiment analysis of Twitter. *Information Processing and Management*, 52(1), 5–19. <http://doi.org/10.1016/j.ipm.2015.01.005>
- Suryani, A. A., Arieshanti, I., Yohanes, B. W., Subair, M., Budiwati, S. D., & Rintyarna, B. S. (2016). Enriching English Into Sundanese and Javanese Translation List Using Pivot Language. In *2016 International Conference on Information, Communication Technology and System (IC)* (pp. 167–171).
- Taboada, M., Brooke, J., & Tofiloski, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(September 2010), 267–307. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/COLI_a_00049