

# Nas\_Jurnal #01 Komparasi Algoritma Klasifikasi Machine Learning

*by* Purwanto Purwanto

---

**Submission date:** 29-Mar-2020 07:10AM (UTC+0700)

**Submission ID:** 1284128546

**File name:** s\_Jurnal\_01\_Komparasi\_Algoritma\_Klasifikasi\_Machine\_Learning.pdf (308.61K)

**Word count:** 3467

**Character count:** 20606

# Komparasi Algoritma Klasifikasi Machine Learning Dan *Feature Selection* pada Analisis Sentimen Review Film

5

Vinita Chandani

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Email: vinita.chandani@gmail.com

5

Romi Satria Wahono, Purwanto

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Email: romi@brainmatics.com, purwanto@dsn.dinus.ac.id

**Abstract:** Analisis sentimen adalah proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks bersifat positif, negatif atau netral. Saat ini, pendapat khalayak umum menjadi sumber yang penting dalam pengambilan keputusan seorang akan suatu produk. Algoritma klasifikasi seperti Naïve Bayes (NB), Support Vector Machine (SVM), dan Artificial Neural Network (ANN) diusulkan oleh banyak peneliti untuk digunakan pada analisis sentimen review film. Namun, klasifikasi sentimen teks mempunyai masalah pada banyaknya atribut yang digunakan pada sebuah dataset. *Feature selection* dapat digunakan untuk mengurangi atribut yang kurang relevan pada dataset. Beberapa algoritma *feature selection* yang digunakan adalah information gain, chi square, forward selection dan backward elimination. Hasil komparasi algoritma, SVM mendapatkan hasil yang terbaik dengan *accuracy* 81.10% dan AUC 0.904. Hasil dari komparasi *feature selection*, information gain mendapatkan hasil yang paling baik dengan *average accuracy* 84.57% dan *average AUC* 0.899. Hasil integrasi algoritma klasifikasi terbaik dan algoritma *feature selection* terbaik menghasilkan *accuracy* 81.50% dan AUC 0.929. Hasil ini mengalami kenaikan jika dibandingkan hasil eksperimen yang menggunakan SVM tanpa *feature selection*. Hasil dari pengujian algoritma *feature selection* terbaik untuk setiap algoritma klasifikasi adalah information gain mendapatkan hasil terbaik untuk digunakan pada algoritma NB, SVM dan ANN.

**Keywords:** analisis sentimen, klasifikasi, feature selection, support vector machine, artificial neural network, naïve bayes, information gain, chi square, forward selection, backward eliminations

## 1 PENDAHULUAN

Analisis sentimen adalah proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks (dokumen, kalimat, paragraf, dll) bersifat positif, negatif atau netral (Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013). Analisis sentimen merupakan bidang penelitian yang cukup populer, karena dapat memberikan keuntungan untuk berbagai aspek, mulai dari prediksi penjualan (Yang Liu, Huang, An, & Yu, 2007), politik (Park, Ko, Kim, Liu, & Song, 2011), dan pengambilan keputusan para investor (Dergiades, 2012).

Saat ini, pendapat khalayak umum telah menjadi salah satu sumber yang begitu penting dalam berbagai produk di jejaring sosial (C.-L. Liu, Hsiao, Lee, Lu, & Jou, 2012). Demikian juga dalam industri film (Tsou & Ma, 2011). Popularitas internet mendorong orang untuk mencari pendapat pengguna dari internet sebelum membeli produk atau melihat situs film (C.-L. Liu et al., 2012). Pendapat orang-orang dapat mengurangi

ketidakpastian terhadap suatu produk tertentu dan membantu konsumen menyimpulkan kualitas suatu produk tertentu (Koh, Hu, & Clemons, 2010).

Banyak situs yang menyediakan review tentang suatu produk yang dapat mencerminkan pendapat pengguna (C.-L. Liu et al., 2012). Salah satu contohnya adalah situs *Internet Movie Database* (IMDb). IMDb adalah situs yang berhubungan dengan film dan produksi film. Informasi yang diberikan IMDb sangat lengkap. Siapa saja aktor/aktris yang main di film itu, sinopsis singkat dari film, link untuk trailer film, tanggal rilis untuk beberapa negara dan review dari user-user yang lain. Ketika seseorang ingin membeli atau menonton suatu film, komentar-komentar orang lain dan peringkat film biasanya mempengaruhi perilaku pembelian mereka.

Algoritma klasifikasi sen seperti naïve bayes (NB) (Kang, Yoo, & Han, 2012), artificial neural network (ANN) (Moraes, Valiati, & Gavião Neto, 2013) (Zhu, Xu, & Wang, 2010), support vector machine (SVM) (Moraes et al., 2013) (S Tan & Zhang, 2008) diusulkan oleh banyak peneliti (Koncz & Alic, 2011) untuk analisis sentimen *review restaurant* (Kang et al., 2012), dokumen (Moraes et al., 2013) (S Tan & Zhang, 2008), dan teks (Zhu et al., 2010). ANN mempunyai kelebihan dalam hal kemampuan untuk generalisasi, yang bergantung pada seberapa baik ANN meminimalkan resiko empiris namun ANN mempunyai kelemahan dimana menggunakan data pelatihan cukup besar (Vapnik, 1999). SVM mempunyai kelebihan yaitu bisa diterapkan untuk data yang berdimensi tinggi, tetapi SVM sulit untuk digunakan untuk data dengan jumlah yang besar (Nugroho, Witarto, & Handoko, 2003). NB mempunyai kelebihan mudah diimplementasikan, performance NB lebih baik. Pengklasifikasian pada NB didasarkan pada probabilitas bersyarat dari fitur salah satu kelas setelah fitur seleksi menggunakan algoritma yang ada (W. Zhang & Gao, 2011).

Beberapa peneliti telah melakukan komparasi menggunakan beberapa algoritma pada beberapa dataset. Penelitian yang dilakukan oleh B. Pang et al (Pang, Lee, Rd, & Jose, 2002) membandingkan algoritma NB, maximum entropy dan SVM didapatkan hasil yang terbaik adalah SVM. Rodrigo Moraes et al (Moraes et al., 2013) membandingkan antara ANN, SVM dan NB. Didapatkan hasil yang terbaik adalah ANN. Ziqiong Zhang et al (Z. Zhang, Ye, Zhang, & Li, 2011) membandingkan antara SVM dan NB dan NB merupakan hasil yang terbaik. Songbo An et al (S Tan & Zhang, 2008) membandingkan NB, centroid classifier, k-nearest neighbor (KNN), winnow classifier dan SVM merupakan hasil yang terbaik. Datas yang digunakan dalam penelitian di atas berbeda-beda. Penelitian yang dilakukan oleh B. Pang et al (Pang & Lee, 2002) menggunakan dataset review film. Rodrigo Moraes et al (Moraes et al., 2013) menggunakan

dataset review film, *Global Positioning System* (GPS), buku dan kamera. Ziqiong Zhang (Z. Zhang et al., 2011) et al menggunakan dataset *review restaurant*, dan Songbo Tan (Songbo Tan & Wang, 2011) et al menggunakan dataset dokumen berbahasa Cina.

Salah satu masalah pada klasifikasi sentimen teks adalah banyaknya atribut yang digunakan pada sebuah dataset (Wang, Li, Song, Wei, & Li, 2011). Pada umumnya, atribut dari klasifikasi sentimen teks sangat besar, dan jika semua atribut tersebut digunakan, maka akan mengurangi kinerja dari *classifier* (Wang, Li, Zhao, & Zhang, 2013). Atribut yang banyak membuat *accuracy* menjadi rendah. Untuk mendapatkan *accuracy* yang lebih baik, atribut yang ada harus dipilih dengan algoritma yang tepat (Xu, Peng, & Cheng, 2012).

*Feature selection* merupakan bagian penting untuk mengoptimalkan kinerja dari *classifier* (Wang et al., 2011). *Feature selection* dapat didasarkan pada pengurangan ruang fitur yang besar, misalnya dengan mengeliminasi atribut yang kurang relevan (Koncz & Paralic, 2011). Penggunaan algoritma *feature selection* yang tepat dapat meningkatkan *accuracy* (Liu et al., 2012) (Forman, 2000). Algoritma *feature selection* dapat dibedakan menjadi dua tipe, yaitu *filter* dan *wrapper* (Yuaning Liu et al., 2011). Contoh dari tipe *filter* adalah *information gain* (IG), *chi-square*, dan *log likelihood ratio*. Contoh dari tipe *wrapper* adalah *forward selection* dan *backward elimination* (Vercellis, 2009). Hasil *precision* dari tipe *wrapper* lebih tinggi daripada tipe *filter*, tetapi hasil ini tercapai dengan tingkat kompleksitas yang besar. Masalah kompleksitas yang tinggi juga dapat menimbulkan masalah (Koncz & Paralic, 2011).

Yang dan Pedersen (Yang & Pedersen, 1997) membandingkan lima algoritma *feature selection* pada klasifikasi dokumen. Lima algoritma tersebut adalah *document frequency*, IG, *chi-square*, *term strength* dan *mutual information*. Hasil penelitian mereka menunjukkan bahwa IG dan *chi-square* paling efisien. Forman (Forman, 2000) membandingkan 12 algoritma *feature selection* pada 229 klasifikasi teks menjadi dua kategori. Hasil penelitian menunjukkan IG dan *chi-square* mendapatkan hasil yang lebih baik dibandingkan metode *Bi-Normal Separation* yang diusulkan peneliti. Tan dan Zang (S Tan & Zhang, 2008) menggunakan algoritma *feature selection* untuk analisis sentimen dokumen berbahasa Cina. Hasil yang didapat IG mendapatkan yang paling baik.

Dari semua hasil penelitian yang sudah dilakukan belum ditemukan model yang paling tepat untuk analisis sentimen. Maka dari itu penulis akan melakukan komparasi terhadap beberapa algoritma klasifikasi (NB, SVM dan ANN), komparasi terhadap beberapa algoritma *feature selection* (IG, *chi-square*, *forward selection*, *backward elimination*) dan melakukan integrasi dari hasil komparasi algoritma klasifikasi dan algoritma *feature selection* yang terbaik pada dataset review film.

## 2 PENELITIAN TERKAIT

Salah satu masalah pada klasifikasi sentiment teks adalah data yang berdimensi tinggi sehingga menyebabkan banyaknya atribut yang kurang relevan. Jika semua atribut tersebut digunakan, maka akan mengurangi kinerja dari sebuah *classifier* (Wang et al., 2013). Atribut yang banyak membuat *accuracy* menjadi rendah. Untuk mendapatkan *accuracy* yang lebih baik, atribut yang ada harus dipilih dengan algoritma yang tepat (Xu et al., 2012). *Feature selection* merupakan

bagian penting untuk mengoptimalkan kinerja dari *classifier* (Wang et al., 2011). *Feature selection* dapat digunakan untuk mengeliminasi atribut yang kurang relevan (Koncz & Paralic, 2011).

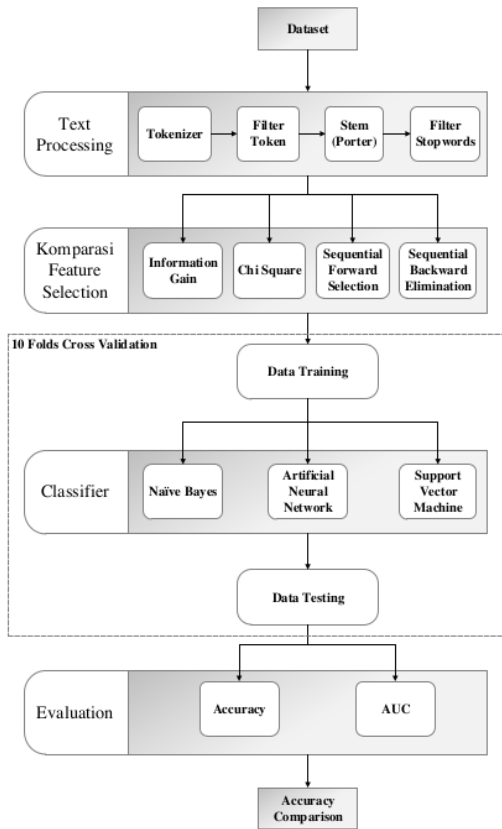
Beberapa peneliti telah mengkomparasi beberapa algoritma klasifikasi dan algoritma *feature selection* untuk mendapatkan hasil yang terbaik. Penelitian yang dilakukan oleh Peter Koncz dan Jan Paralic (Koncz & Paralic, 2011) menggunakan SVM untuk algoritma klasifikasinya dan algoritma *feature selection* n-grams+ *document frequency* dibandingkan dengan Information Gain (IG). Hasil yang didapatkan IG lebih baik daripada algoritma yang diusulkan. Rodrigo Moraes, Joao Francisco Valiati, Wilson P (Moraes et al., 2013) mengkomparasi algoritma klasifikasi SVM, Naïve Bayes (NB) dan Artificial Neural Network (ANN). *Feature selection* yang digunakan adalah *expert knowledge*, *minimum frequency*, IG, *chi-square*. Hasil yang terbaik untuk algoritma klasifikasi adalah ANN dan untuk *feature selection* terbaik adalah IG. Zhu Jian, Xu Chen dan Wang Han Shi (Zhu et al., 2010) mengkomparasi algoritma klasifikasi *individual model (i-model)* berbasis ANN dibandingkan dengan hidden markov model dan SVM. *Feature selection* yang digunakan adalah *odd ratio*. Hasil algoritma klasifikasi yang terbaik adalah *i-model based on ANN*. Songbo Tan dan Jin Zhang (S Tan & Zhang, 2008) mengkomparasi lima algoritma klasifikasi (*centroid classifier*, *K-nearest neighbor*, *winnow classifier*, NB dan SVM), empat algoritma *feature selection* (*Mutual Information*, IG, *chi-square* dan *Document Frequency*). Hasil eksperimen menunjukkan bahwa IG mendapatkan hasil yang terbaik untuk *feature selection* dan algoritma SVM mendapatkan hasil yang terbaik untuk klasifikasi sentimen.

## 3 METODE YANG DIUSULKAN

Peneliti mengusulkan untuk mengkomparasi tiga algoritma klasifikasi (SVM, NB dan ANN) dan mengkomparasi empat algoritma *feature selection* (IG, *Chi Square*, *Forward Selection* dan *Backward Elimination*). Gambar 1 menunjukkan komparasi algoritma klasifikasi dan *feature selection* yang diusulkan.

Sebelum dilakukan komparasi, dataset dilakukan *text processing* terlebih dahulu. *Text processing* bertujuan untuk mempersiapkan dokumen teks yang tidak terstruktur menjadi data terstruktur yang siap digunakan untuk proses selanjutnya. Tahapan *text processing* meliputi:

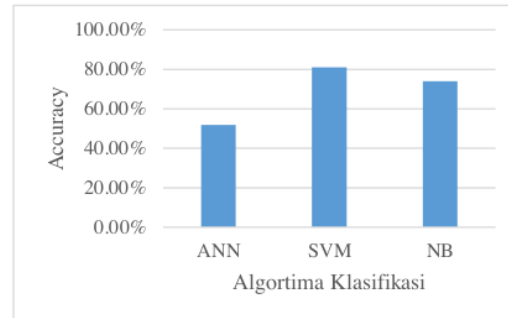
1. Tokenize merupakan proses untuk memisah-misahkan kata. Potongan kata tersebut disebut dengan token atau *term* (Manning, Raghavan, & Schütze, n.d.).
2. Filter Token merupakan proses mengambil kata-kata penting dari hasil token (Langgeni, Baizal, & W, 2010).
3. Stem yaitu proses perubahan bentuk kata menjadi kata dasar. Metode perubahan bentuk kata menjadi kata dasar ini menyesuaikan struktur bahasa yang digunakan dalam proses stemming (Langgeni et al., 2010).
4. Filter stopwords adalah proses menghilangkan kata-kata yang sering muncul namun tidak memiliki pengaruh apapun dalam ekstraksi sentimen suatu review. Kata yang termasuk seperti kata penunjuk waktu, kata tanya (Langgeni et al., 2010).



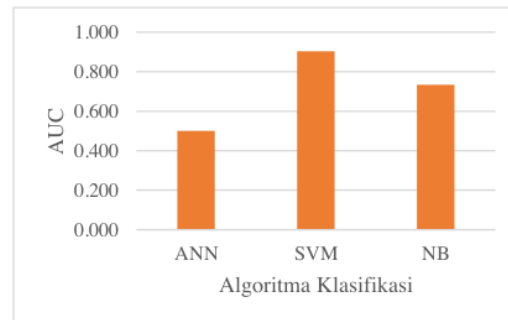
Gambar 1. Komparasi Algoritma Klasifikasi dan *Feature Selection*

Tabel 1. Komparasi Accuracy dan AUC Algoritma Klasifikasi

	Accuracy	AUC
ANN	51.80%	0.500
SVM	81.10%	0.904
NB	74.00%	0.734



Gambar 2. Komparasi accuracy algoritma klasifikasi



Gambar 3. Komparasi AUC algoritma klasifikasi

4 HASIL PENELITIAN

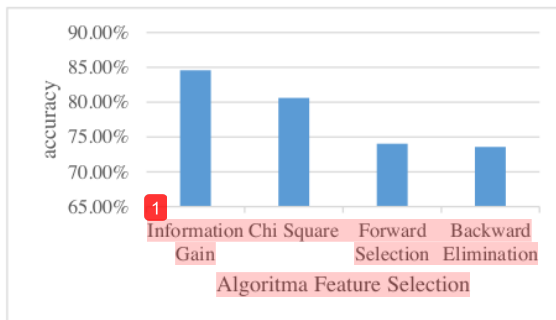
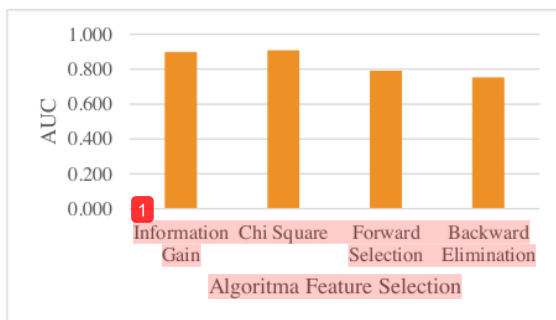
Penelitian yang dilakukan menggunakan komputer dengan spesifikasi CPU Intel Core i5 1.6GHz, RAM 8GB, dan sistem operasi Microsoft Windows 7 Professional 64-bit. Aplikasi yang digunakan adalah 20 ipidMiner 5.2. Data penelitian ini menggunakan *Data Movie Review Polarity Dataset V2.0* (Pang & Lee, 2002) yang diperoleh dari data movie review 12g digunakan oleh Pang and Lee. Data ini dapat diambil di situs <http://www.cs.cornell.edu/people/pabo/movie-review-data/>. Data ini diambil dari situs IMDb. Data yang digunakan dalam penelitian terdiri dari 1000 review film, berisi 500 review positif dan 500 review negatif.

Tabel 5 merupakan rangkuman hasil komparasi algoritma klasifikasi. Berdasarkan Tabel 1, Gambar 2 dan Gambar 3 didapat hasil terbaik adalah SVM dengan *accuracy* = 81.10% dan *AUC* = 0.904. Hal ini mengkonfirmasi pada penelitian yang dilakukan oleh Songbo Tan (S Tan & Zhang, 2008) dalam mengkomparasi algoritma klasifikasi, dan SVM mendapatkan nilai yang paling baik. Klasifikasi pada analisis sentimen sangat tergantung pada data yang diuji. Untuk pengujian data IMDB review film, SVM merupakan algoritma yang paling baik.

SVM menghasilkan nilai *accuracy* dan *AUC* terbaik dibanding ANN dan NB. Hal ini mengkonfirmasi pada penelitian yang dilakukan oleh Songbo Tan (S Tan & Zhang, 2008) dalam mengkomparasi algoritma klasifikasi, dan SVM mendapatkan nilai yang paling baik. Klasifikasi pada analisis sentimen sangat tergantung pada data yang diuji. Untuk pengujian data IMDB review film, SVM merupakan algoritma yang paling baik.

Tabel 2. Komparasi Accuracy dan AUC Algoritma *Feature Selection*

	Information Gain		Chi Square		Forward Selection		Backward Elimination	
	Top K (K=200)		Top K (K=100)					
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
ANN	91.40%	0.914	79.60%	0.900	75.50%	0.781	70.20%	0.724
SVM	81.50%	0.929	80.80%	0.853	67.67%	0.698	79.25%	0.844
NB	80.80%	0.853	80.30%	0.867	79.00%	0.807	71.25%	0.689
AVERAGE	84.57%	0.899	80.23%	0.873	74.06%	0.762	73.57%	0.752

Gambar 4. Grafik Komparasi Accuracy Algoritma *Feature Selection*Gambar 5. Grafik Komparasi AUC Algoritma *Feature Selection*

Tabel 2 merupakan tabel komparasi *feature selection* terbaik. Data dari Tabel 2 diambil berdasarkan *average* (rata-rata) dari masing-masing parameter algoritma *feature selection*. Dari hasil *average* tersebut, diambil nilai *average* yang paling baik, dan kemudian dirangkumkan seperti pada Tabel 2. Berdasarkan Tabel 2 didapatkan hasil algoritma *feature selection* terbaik adalah *information gain*. Hal ini mengkonfirmasi pada penelitian yang dilakukan oleh Peter Koncz (Koncz & Paralic, 2011), Rodrigo Moraes (Moraes et al., 2013), dan Songbo Tan (S Tan & Zhang, 2008) yang juga menghasilkan *information gain* sebagai algoritma *feature selection* yang terbaik.

## 5 KESIMPULAN

Hasil dari komparasi algoritma klasifikasi antara Support Vector Machine (SVM), Naïve Bayes (NB) dan Artificial Neural Network (ANN) didapatkan SVM dengan hasil terbaik dengan nilai *accuracy* = 81.10% dan nilai AUC = 0.904. Hasil

dari komparasi algoritma *feature selection* antara *information gain*, *chi square*, *forward selection*, *backward elimination* didapatkan *information gain* pada parameter top k dengan nilai k = 200 sebagai hasil terbaik, dengan nilai *accuracy average* adalah 84.57% dan nilai AUC = 0.899.

## REFERENCES

- Dergiades, T. (2012). Do investors' sentiment dynamics affect stock returns? Evidence from the US economy. *Economics Letters*, 116(3), 404–407. doi:10.1016/j.econlet.2012.04.018
- Forman, G. (2000). An Extensive Empirical Study of Feature Selection Metrics for Text Classification. *Journal of Machine Learning Research*, 3, 1289–1305. doi:10.1162/153244303322753670
- Kang, H., Yoo, S. J., & Han, D. (2012). Senti lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5), 6000–6010. doi:10.1016/j.eswa.2011.11.107
- Koh, N. S., Hu, N., & Clemons, E. K. (2010). Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications*, 9(5), 374–385. doi:10.1016/j.elerap.2010.04.001
- Koncz, P., & Paralic, J. (2011). An approach to feature selection for sentiment analysis. In *2011 15th IEEE International Conference on Intelligent Engineering Systems* (pp. 357–362). IEEE. doi:10.1109/INES.2011.5954773
- Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, 40(10), 4065–4074. doi:10.1016/j.eswa.2013.01.001
- Langgeni, D. P., Baizal, Z. K. A., & W, Y. F. A. (2010). Clustering Artikel Berita Berbahasa Indonesia, 2010(semnasIF), 1–10.
- Liu, C.-L., Hsiao, W.-H., Lee, C.-H., Lu, G.-C., & Jou, E. (2012). Movie Rating and Review Summarization in Mobile Environment. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(3), 397–407. doi:10.1109/TSMCC.2011.2136334
- Liu, Y., Huang, X., An, A., & Yu, X. (2007). ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07* (p. 607). New York, New York, USA: ACM Press. doi:10.1145/1277741.1277845
- Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., & Wang, S. (2011). An Improved Particle Swarm Optimization for Feature Selection. *Journal of Bionic Engineering*, 8(2), 191–200. doi:10.1016/S1672-6529(11)60020-6
- Manning, C. D., Raghavan, P., & Schütze, H. (n.d.). Introduction to Information Retrieval.
- Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document Level Sentiment Classification: an Empirical Comparison

between SVM and ANN. *Expert Systems with Applications*, 40(2), 621–633. doi:10.1016/j.eswa.2012.07.059

- Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika. *IlmuKomputer.Com*.
- Pang, B., & Lee, L. (2002). A Sentimental Education : Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Association for Computational Linguistics*.
- Pang, B., Lee, L., Rd, H., & Jose, S. (2002). Thumbs up ? Sentiment Classification using Machine Learning Techniques. *Association for Computational Linguistics*, 10(July), 79–86.
- Park, S., Ko, M., Kim, J., Liu, Y., & Song, J. (2011). The Politics of Comments : Predicting Political Orientation of News Stories with Commenters' Sentiment Patterns.
- Tan, S., & Wang, Y. (2011). Weighted SCL model for adaptation of sentiment classification. *Expert Systems with Applications*, 38(8), 10524–10531. doi:10.1016/j.eswa.2011.02.106
- Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications*, 34(4), 2622–2629. doi:10.1016/j.eswa.2007.05.028
- Tsou, B. K., & Ma, M. (2011). Aspect Based Opinion Polling from Customer Reviews. *IEEE Transactions on Affective Computing*, 2(1), 37–49. doi:10.1109/T-AFFC.2011.2
- Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, 10(5), 988–99. doi:10.1109/72.788640
- Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. John Wiley and Sons.
- Wang, S., Li, D., Song, X., Wei, Y., & Li, H. (2011). A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. *Expert Systems with Applications*, 38(7), 8696–8702. doi:10.1016/j.eswa.2011.01.077
- Wang, S., Li, D., Zhao, L., & Zhang, J. (2013). Sample cutting method for imbalanced text sentiment classification based on BRC. *Knowledge-Based Systems*, 37, 451–461. doi:10.1016/j.knosys.2012.09.003
- Xu, T., Peng, Q., & Cheng, Y. (2012). Identifying the semantic orientation of terms using S-HAL for sentiment analysis. *Knowledge-Based Systems*, 35, 279–289. doi:10.1016/j.knosys.2012.04.011
- Yang, Y., & Pedersen, J. O. (1997). A Comparative Study on Feature Selection in Text Categorization. *Proceedings of the Fourteenth International Conference on Machine Learning*, 20(15), 412–420.
- Zhang, W., & Gao, F. (2011). An Improvement to Naive Bayes for Text Classification. *Advanced in Control Engineering and Information Science*, 15, 2160–2164. doi:10.1016/j.proeng.2011.08.404
- Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 38(6), 7674–7682. doi:10.1016/j.eswa.2010.12.147
- Zhu, J., Xu, C., & Wang, H. (2010). Sentiment classification using the theory of ANNs. *The Journal of China Universities of Posts and Telecommunications*, 17(July), 58–62. doi:10.1016/S1005-8885(09)60606-3

## BIOGRAFI PENULIS



**Vinita Chandani.** Lahir pada tanggal 11 November 1990 di Tegal, Jawa Tengah. Memperoleh gelar Sarjana Komputer (S.Kom) dari fakultas Teknik Informatika, Universitas Aki Semarang pada tahun 2011. Serta memperoleh gelar M.Kom dari Fakultas Ilmu Komputer, Universitas Dian Nuswantoro pada tahun 2014.



**Romi Satria Wahono.** Memperoleh Gelar B.Eng dan M.Eng pada bidang ilmu komputer di Saitama University, Japan, dan Ph.D pada bidang software engineering di Universiti Teknikal Malaysia Melaka. Menjadi pengajar dan peneliti di Fakultas Ilmu Komputer, Universitas Dian Nuswantoro. Merupakan pendiri dan CEO PT Brainmatics, sebuah perusahaan yang bergerak di bidang pengembangan software. Minat penelitian pada bidang software engineering dan machine learning. Profesional member dari asosiasi ilmiah ACM, PMI dan IEEE Computer Society.



**Purwanto.** Menyelesaikan pendidikan S1 di Universitas Diponegoro Semarang, S2 di STMIK Benarif Indonesia dan S3 di Universitas Multimedia Malaysia. Saat ini menjadi dosen pascasarjana Magister Teknik Informatika di Universitas Dian Nuswantoro. Minat penelitian saat ini adalah data mining, machine learning, soft computing, artificial intelligence, decision support system.

# Nas\_Jurnal #01 Komparasi Algoritma Klasifikasi Machine Learning

## ORIGINALITY REPORT

18%

SIMILARITY INDEX

9%

INTERNET SOURCES

7%

PUBLICATIONS

9%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://www.academia.edu">www.academia.edu</a> Internet Source	2%
2	Submitted to Sriwijaya University Student Paper	2%
3	Yumi Novita Dewi, Dwiza Riana, Teddy Mantoro. "Improving Naïve Bayes performance in single image pap smear using weighted principal component analysis (WPCA)", 2017 International Conference on Computing, Engineering, and Design (ICCED), 2017 Publication	2%
4	<a href="http://jurnal.fikom.umi.ac.id">jurnal.fikom.umi.ac.id</a> Internet Source	2%
5	Submitted to Universitas Brawijaya Student Paper	1%
6	<a href="http://research.pps.dinus.ac.id">research.pps.dinus.ac.id</a> Internet Source	1%
7	Submitted to Universitas Mercu Buana	

1%

8

[www.dinus.ac.id](http://www.dinus.ac.id)

Internet Source

1%

9

Peter Koncz, Jan Paralic. "An approach to feature selection for sentiment analysis", 2011  
15th IEEE International Conference on Intelligent Engineering Systems, 2011

Publication

1%

10

[www.coursehero.com](http://www.coursehero.com)

Internet Source

1%

11

Valonia Inge Santoso, Gloria Virginia, Yuan Lukito. "PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE",  
Jurnal Transformatika, 2017

Publication

1%

12

[ejournal.unib.ac.id](http://ejournal.unib.ac.id)

Internet Source

1%

13

Piryani, R., D. Madhavi, and V.K. Singh. "Analytical mapping of opinion mining and sentiment analysis research during 2000-2015", Information Processing & Management, 2016.

Publication

1%

Riska Aryanti, Atang Saepudin, Eka Fitriani,



14

Rifky Permana, Dede Firmansyah Saefudin. "Komparasi Algoritma Naive Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Pengguna Busway", Jurnal Teknik Komputer, 2019

Publication

1%

15

kt.ijs.si

Internet Source

<1%

16

Sharma, Anuj, and Shubhamoy Dey. "A document-level sentiment analysis approach using artificial neural network and sentiment lexicons", ACM SIGAPP Applied Computing Review, 2012.

Publication

<1%

17

"Market Analysis Using Computational Intelligence: An Application for GSM Operators Based on Twitter Comments", Intelligent Systems Reference Library, 2015.

Publication

<1%

18

Haiyun Peng, Erik Cambria, Amir Hussain. "A Review of Sentiment Analysis Research in Chinese Language", Cognitive Computation, 2017

Publication

<1%

19

Submitted to Forum Komunikasi Perpustakaan Perguruan Tinggi Kristen Indonesia (FKPPTKI)

Student Paper

<1%

20	aclweb.org Internet Source	<1%
21	Submitted to Universitas Merdeka Malang Student Paper	<1%
22	www.jatit.org Internet Source	<1%
23	www.duo.uio.no Internet Source	<1%
24	Submitted to Politeknik Negeri Bandung Student Paper	<1%
25	Gholamian, Mohammad Reza, and Samaneh Mazidi Sharaf Abadi. "Resource evaluation: a novel application of opinion mining for competitive intelligence", International Journal of Business Information Systems, 2014. Publication	<1%

Exclude quotes Off  
Exclude bibliography On

Exclude matches < 5 words