ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *NAÏVE BAYES* PADA ANALISIS SENTIMEN SOSIAL MEDIA TWITTER

(Studi Kasus Kuliah Luring Setelah Pandemi Covid-19)

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana Matematika

Oleh:

Zahra Agusfiyanti Nurlaila

NIM 1905657

PROGRAM STUDI MATEMATIKA

DEPARTEMEN PENDIDIKAN MATEMATIKA

FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS PENDIDIKAN INDONESIA

2023

LEMBAR HAK CIPTA

ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *NAÏVE BAYES* PADA ANALISIS SENTIMEN SOSIAL MEDIA TWITTER
(Studi Kasus Kuliah Luring Setelah Pandemi Covid-19)

Oleh:
Zahra Agusfiyanti Nurlaila
NIM 1905657

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Matematika pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

LEMBAR PENGESAHAN

ZAHRA AGUSFIYANTI NURLAILA

ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *NAÏVE BAYES* PADA ANALISIS SENTIMEN SOSIAL MEDIA TWITTER
(Studi Kasus Kuliah Luring Setelah Pandemi Covid-19)

Disetujui dan disahkan oleh :

Pembimbing I

Dr. Dadan Dasari, M.Si.
NIP. 196407171991021001

Pembimbing II

Dr. Lukman, S.Si., M.Si.
NIP. 196801281994021001

Mengetahui,

Ketua Program Studi Matematika

Dr. Kartika Yulianti, S.Pd., M.Si.
NIP. 199207282005012001

ABSTRAK

Setelah melewati masa pandemi Covid-19, berbagai sektor dalam kehidupan manusia banyak mengalami perubahan, termasuk dalam kegiatan pendidikan. Hal tersebut membuat berbagai pendapat atau tanggapan mengenai kegiatan pendidikan banyak dituangkan dalam media sosial. Untuk mengetahui sentimen tanggapan masyarakat tersebut perlu dilakukan analisis sentimen dengan algoritma *machine learning*. Metode yang digunakan dalam penelitian ini adalah *Support Vector Machine*, *Naïve Bayes Classifier*, dan *Lexicon Based*. SVM digunakan untuk proses klasifikasi analisis sentimen dan mencari nilai akurasi terbaik menggunakan *kernel linear*. *Naïve Bayes* digunakan untuk proses klasifikasi analisis sentimen dan sebagai metode perbandingan terhadap SVM. *Lexicon Based* digunakan untuk menentukan kelas sentimen positif, netral dan negatif pada data. Hasil penilaian dari 3522 data *tweet* diperoleh 766 *tweet* (21.7%) positif, 254 tweet netral (7.2%), dan 2502 tweet negatif (71%). Metode klasifikasi SVM memiliki tingkat akurasi sebesar 83% , presisi sebesar 78%, dan *recall* sebesar 55. Sedangkan metode *Naïve Bayes* memiliki tingkat akurasi sebesar 73%, nilai presisi 58% dan *recall* sebesar 34%. Berdasarkan hasil analisis sentimen ini, dapat disimpulkan bahwa performa metode SVM lebih baik dalam mengklasifikasi data dibandingkan *Naïve Bayes*.

Kata Kunci: Analisis Sentimen, Komentar, Opini, *Twitter*, *Support Vector Machine*, SVM, *Naïve Bayes*, *Lexicon Based*, Kuliah, Luring

ABSTRACT

After going through the Covid-19 pandemic, various sectors in human life experienced many changes, including in educational activities. This causes various opinions or responses regarding educational activities to be poured on social media. To find out the sentiment of the public's response, it is necessary to carry out sentiment analysis with a machine learning algorithm. The methods used in this research are Support Vector Machine, Naïve Bayes Classifier, and Lexicon Based. SVM is used for the sentiment analysis classification process and to find the best accuracy value using a linear kernel. Naïve Bayes is used for the sentiment analysis classification process and as a comparison method against SVM. Lexicon Based is used to determine the positive, neutral and negative sentiment classes in the data. The results of the assessment of 3522 tweet data obtained 766 positive tweets (21.7%), 254 neutral tweets (7.2%), and 2502 negative tweets (71%). The SVM classification method has an accuracy rate of 83%, a precision of 78%, and a recall of 55. Meanwhile the Naïve Bayes method has an accuracy rate of 73%, a precision value of 58% and a recall of 34%. Based on the results of this sentiment analysis, it can be concluded that the performance of the SVM method is better in classifying data than Naïve Bayes.

***Keywords**: Sentiment Analysis, Comments, Opinion, Twitter, Support Vector Machine, SVM, Naïve Bayes, Lexicon Based, Lecture, Offline*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR PUSTAKA

Admin LP2M. (2022, February 21). *Analisis Sentimen (Sentiment Analysis) : Definisi, Tipe dan Cara Kerjanya*. Lembaga Penelitian Dan Pengabdian Masyarakat (LP2M) Universitas Medan Area.

Alencar, P., & Cowan, D. (2018). *The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review Ivens Portugal*.

Alfi, M., Reynaldhi, R., & Sibaroni, Y. (2015). *Analisis Sentimen Review Film pada Twitter menggunakan Metode Klasifikasi Hybrid SVM, Naïve Bayes, dan Decision Tree*.

Arini, A.-, Wardhani, L. K., & Octaviano, D.-. (2020). Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden. *KILAT*, *9*(1), 103–114. https://doi.org/10.33322/kilat.v9i1.878

Camelia, D. (2023, February 16). *Apa Itu Analisis Sentimen : Pengertian, Tipe, Dan Cara Kerjanya*. Kazee.Id.

Darwis, D., Shintya Pratiwi, E., Ferico, A., & Pasaribu, O. (2020). PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA. In *Jurnal Ilmiah Edutic* (Vol. 7, Issue 1).

Esti. (2023, August 9). *Mengenal Apa Itu Analisis Sentimen, Tipe dan Cara Kerjanya*. Mekari Qontak.

Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, *8*(1), 11–16. https://doi.org/10.34148/teknika.v8i1.95

Fauzi, A. (2020). Studi Perbandingan Metode Analisis Naive Bayes Classifier dengan Support Vector Machine untuk Analisis Sentimen (Studi Kasus: Tweet Berbahasa Indonesia tentang Covid-19). Yogyakarta.

Fauzi, R. R. (2022) Analisis Sentimen Dampak Ekonomi Masyarakat Indonesia Akibat Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier, Support Vector Machine Dan Lexicon Based. Jakarta.

Feinerer, I., Hornik, K., & Meyer, D. (2008). *Journal of Statistical Software Text Mining Infrastructure in R*. http://www.jstatsoft.org/

Feldman, R., & Sanger, J. (2007). *The text mining handbook : advanced approaches in analyzing unstructured data*. Cambridge University Press.

Ghahramani, Saeed. (2005). *Fundamentals of probability, with stochastic processes*.

Han, J., & Kamber, M. (2006). *Data Mining Concept and Techniques*.

Hendry, J. (2012, Mei 30). Prior Probability Vs Posterior Probability. Retrieved 8 5, 2023 from scribd.com: https://www.scribd.com/doc/95263438/Prior

Kashina, M., Lenivtceva, I. D., & Kopanitsa, G. D. (2020). Preprocessing of unstructured medical data: The impact of each preprocessing stage on classification. *Procedia Computer Science*, *178*, 284–290. https://doi.org/10.1016/j.procs.2020.11.030

Kurniasari, D. (2021, May 6). *Algoritma Supervised Learning vs Unsupervised Learning Cari Tahu Perbedaannya Disini*. DQLAB AI Powered Learning.

Kwok, L., & Yu, B. (2013). Spreading Social Media Messages on Facebook An Analysis of Restaurant Business-to-Consumer Communications. *Cornell Hospitality Quarterly*, *54*, 84–94. https://doi.org/10.1177/1938965512458360

LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. Nature 521, 436–444.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Mukhtar, N., Khan, M. A., & Chiragh, N. (2018). Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains. *Telematics and Informatics*, *35*(8), 2173–2183. https://doi.org/https://doi.org/10.1016/j.tele.2018.08.003

Nasteski, V. (2017). An overview of the supervised machine learning methods. *HORIZONS.B*, *4*, 51–62. https://doi.org/10.20544/horizons.b.04.1.17.p05

Nindito, H. (2016). TEORI TEXT MINING DAN WEB MINING. https://sis.binus.ac.id/2016/12/15/teori-text-mining-dan-web-mining/

Novakovic, J., Veljovic, A., Ilić, S., Papic, Ž. M., & Milica, T. (2017). Evaluation of Classification Models in Machine Learning. *Theory and Applications of Mathematics & Computer Science*, *7*, 39–46. https://api.semanticscholar.org/CorpusID:1586327

Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). *Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika 1*. http://asnugroho.net

P, R. D. L., Fatichah, C., & Purwitasari, D. (2017). Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision Tree, Random Forest, dan SVM. *Jurnal Teknik ITS*, *6*(1), 153–158.

Pujianto, U., & Ristanti, P. Y. (n.d.). *Jurusan Teknik Elektro, Universitas Negeri Malang, Indonesia | Maret 2019 U. Pujianto, Putri Yuni Ristanti | Perbandingan kinerja metode C4.5 dan Naive Bayes dalam klasifikasi …* (Vol. 29). http://journal2.um.ac.id/index.php/tekno

Ramos, J. (2003). *Using TF-IDF to determine word relevance in document queries*.

Rania, D. (2022, June 28). *Apa Itu Big Data? Arti, Konsep, Contoh, dan Manfaatnya*. Rumah Web.

Rish, I. (2006). *An empirical study of the naive Bayes classifier*.

Robertson, S. (2004). Understanding Inverse Document Frequency: On Theoretical Arguments for IDF. *Journal of Documentation - J DOC*, *60*, 503–520. https://doi.org/10.1108/00220410410560582

S. N. &. F. K. Asiyah, "Klasifikasi Berita Online Menggunakan Metode Support Vector Machine dan K-Nearest Neighbor," Jurnal Sains & Seni, pp. 317-322, 2016.

Sammut, C., & Webb, G. (2010). Encyclopedia of Machine Learning. In *Encyclopedia of Machine Learning: , ISBN 978-0-387-30768-8. Springer Science+Business Media, LLC, 2010*. https://doi.org/10.1007/978-0-387-30164-8

Schmidhuber, J. (2014). *Deep Learning in Neural Networks: An Overview*. https://doi.org/10.1016/j.neunet.2014.09.003

Sulardi, N., & Witanti, A. (2020). Sistem Pakar Untuk Diagnosis Penyakit Anemia Menggunakan Teorema Bayes. *Jurnal Teknik Informatika*, *1*(1), 19–24. https://doi.org/10.20884/1.jutif.2020.1.1.12

Tiara, Sabariah, M. K., & Effendy, V. (2015). Sentiment analysis on Twitter using the combination of lexicon-based and support vector machine for assessing the performance of a television program. *2015 3rd International Conference on Information and Communication Technology (ICoICT)*, 386–390. https://api.semanticscholar.org/CorpusID:18189720

Wahyudi, R., Kusumawardhana, G., Purwokerto, A., Letjend, J., Soemarto, P., Purwanegara, K., Purwokerto, T., & Banyumas, K. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *JURNAL INFORMATIKA*, *8*(2). http://ejournal.bsi.ac.id/ejurnal/index.php/ji

Winiarti, S. (2013). Pemanfatan dalam penentuan penyakit tht. Journal of Chemical Information and Modeling, 53(9), 1689–1699.

Wira, J., & Putra, G. (2020) *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning Edisi 1.4*.

Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information Sciences*, *507*, 772–794. https://doi.org/https://doi.org/10.1016/j.ins.2019.06.064

Xu, J., Zhang, Y., Wu, Y., & Wang, J. (2015). *Citation Sentiment Analysis in Clinical Trial Papers Article in AMIA*. http://www.ncbi.nlm.nih.gov/pmc/

Zoqi Sarwani, M., & Mahmudy, W. (2015). *Analisis twitter untuk mengetahui karakter seseorang menggunakan algoritma naïve bayess classifier*.

Zukhrufillah, I. (2018). *Gejala Media Sosial Twitter Sebagai Media Sosial Alternatif* (Vol. 1, Issue 2). https://id.wikipedia.org/wiki/Twitter