DAFTAR PUSTAKA

[1]   A. S. Nugroho, A. B. Witarto, and D. Handoko, "Support Vector Machine, Teori dan Aplikasinya dalam Bioinformatika," 2003. [Online]. Available: http://asnugroho.net/papers/ikcsvm.pdf. [Accessed: 08-Dec-2017].

[2]   T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated Hate Speech Detection and the Problem of Offensive Language," *Proc. 11th Int. AAAI Conf. Web Soc. Media*, no. Icwsm, pp. 512–515, 2017.

[3]   S. H. Yadav and P. M. Manwatkar, "An approach for offensive text detection and prevention in Social Networks," *ICIIECS 2015 - 2015 IEEE Int. Conf. Innov. Information, Embed. Commun. Syst.*, pp. 3–6, 2015.

[4]   J. Du and Z. A. Yi, "A New Knn Categorization Algorithm for Harmful Information Filtering," in *2012 Fifth International Symposium on Computational Intelligence and Design*, 2012, pp. 489–492.

[5]   N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[6]   N. V Chawla, N. Japkowicz, and P. Drive, "Editorial : Special Issue on Learning from Imbalanced Data Sets," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 1–6, 2004.

[7]   C. Drummond and R. C. Holte, "C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling," *Work. Learn. from Imbalanced Datasets II*, pp. 1–8, 2003.

[8]   A. C. Liu, "The Effect of Oversampling and Undersampling on Classifying Imbalanced Text Datasets," University of Texas at Austin, 2004.

[9]   F. A. Wenando, "KLASIFIKASI TEKS UNTUK MENDETEKSI TINGKAT PEMAHAMAN MAHASISWA DALAM PROSES *PRIOR KNOWLEDGE ACTIVATION,*" Universitas Gadjah Mada, Yogyakarta, Indonesia, 2017.

[10]   A. Estabrooks, T. Jo, and N. Japkowicz, "A Multiple Resampling Method for Learning from Imbalanced Data Sets," *Comput. Intell.*, vol. 20, no. 1, pp. 18–36, Feb. 2004.

[11]   A. Liu, C. Martin, B. La Cour, and J. Ghosh, "Effects of Oversampling Versus Cost-Sensitive Learning for Bayesian and SVM Classifiers," in *Data Mining: Special Issue in Annals of Information Systems*, vol. 8, R. Stahlbock, S. F. Crone, and S. Lessmann, Eds. Springer US, 2010, pp. 159–192.

[12] Z. Waseem and D. Hovy, "Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter," *Proc. NAACL Student Res. Work.*, pp. 88–93, 2016.

[13] G. M. Weiss, "Foundations of Imbalanced Learning," in *Imbalanced Learning*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013, pp. 13–41.

[14] "International Covenant on Civil and Political Rights." [Online]. Available: http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx. [Accessed: 15-Nov-2017].

[15] P. Fortuna, "Automatic detection of hate speech in text : an overview of the topic and dataset annotation with hierarchical classes," 2017.

[16] "Kebijakan perilaku kebencian." [Online]. Available: https://support.twitter.com/articles/20175055. [Accessed: 15-Nov-2017].

[17] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[18] G. M. Weiss and F. J. Provost, "Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction.," *J. Artif. Intell. Res.*, vol. 19, pp. 315–354, 2003.

[19] M. C. Monard and G. E. A. P. A. Batista, "Learning with skewed class distributions," *Adv. Logic, Artif. Intell. Robot.*, pp. 173–180, 2002.

[20] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, 2004.

[21] G. Lemaitre, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *CoRR*, vol. abs/1609.0, pp. 1–5, 2016.

[22] D. Arthur and S. Vassilvitskii, "K-Means++: the Advantages of Careful Seeding," *Proc. eighteenth Annu. ACM-SIAM Symp. Discret. algorithms*, vol. 8, pp. 1027–1025, 2007.

[23] J. Zhang and I. Mani, "kNN Approach to Unbalanced Data Distributions: A Case Study involving Information Extraction," *Work. Learn. from Imbalanced Datasets II ICML Washingt. DC 2003*, pp. 42–48, 2003.

[24] M. R. Smith, T. Martinez, and C. Giraud-Carrier, "An instance level analysis of data complexity," *Mach. Learn.*, vol. 95, no. 2, pp. 225–256, 2014.

[25] G. Lemaitre, "Computer-Aided Diagnosis for Prostate Cancer using Multi-Parametric Magnetic Resonance Imaging," Universitat de Girona, 2016.

[26] F. A. Wenando, T. B. Adji, and I. Ardiyanto, "Text Classification to Detect Student Level of Understanding in Prior Knowledge Activation Process," *Adv. Sci. Lett.*, vol. 23, no. 3, pp. 2285–2287, Mar. 2017.

[27] M. Chau and H. Chen, "A machine learning approach to web page filtering using content and structure analysis," *Decis. Support Syst.*, vol. 44, no. 2, pp. 482–494, 2008.

[28] W. Fan, M. D. Gordon, and P. Pathak, "An integrated two-stage model for intelligent information routing," *Decis. Support Syst.*, vol. 42, no. 1, pp. 362–374, 2006.

[29] T. Joachims, "Text categorization with Support Vector Machines: Learning with many relevant features," in *Machine Learning*, vol. 1398, no. LS-8 Report 23, 1998, pp. 137–142.

[30] A. Sun, E.-P. Lim, and Y. Liu, "On strategies for imbalanced text classification using SVM: A comparative study," *Decis. Support Syst.*, vol. 48, no. 1, pp. 191–201, 2009.

[31] P. Hart, "The condensed nearest neighbor rule (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 515–516, 1968.

[32] D. L. Wilson, "Asymptotic Properties of Nearest Neighbor Rules Using Edited Data," *IEEE Trans. Syst. Man Cybern.*, vol. 2, no. 3, pp. 408–421, 1972.

[33] S. Shekarforoush, R. Green, and R. Dyer, "Classifying commit messages: A case study in resampling techniques," *Proc. Int. Jt. Conf. Neural Networks*, pp. 1273–1280, 2017.

[34] I. Tomek, "An Experiment with the Edited Nearest-Neighbor Rule," *IEEE Trans. Syst. Man. Cybern.*, vol. 6, no. 6, pp. 448–452, 1976.

[35] S. Vluymans, "Instance Selection for Imbalanced Data," Ghent University, 2014.

[36] M. Kubat and S. Matwin, "Addressing the Curse of Imbalanced Training Sets: One Sided Selection," *Icml*, vol. 97, pp. 179–186, 1997.

[37] I. Tomek, "Two Modifications of CNN," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-6, no. 11, pp. 769–772, 1976.

[38] A. Fernández, V. López, M. Galar, M. J. Del Jesus, and F. Herrera, "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches," *Knowledge-Based Syst.*, vol. 42, pp. 97–110, 2013.

[39]   P. Kang, S. Cho, and D. L. MacLachlan, "Improved response modeling based on clustering, under-sampling, and ensemble," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 6738–6753, 2012.

[40]   T. Joachims, *Learning to Classify Text Using Support Vector Machines*. Springer Science & Bussines Media, 2012.

[41]   W. Zhang, T. Yoshida, and X. Tang, "A comparative study of TF*IDF, LSI and multi-words for text classification," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2758–2765, 2011.

[42]   C. Cortes and V. N. Vapnik, "Support-vector networks," *Mach. Learn. J.*, vol. 20, pp. 273–297, 1995.

[43]   V. N. Vapnik, *Statistical Learning Theory*. Wiley-Interscience, 1998.

[44]   V. N. Vapnik, *Estimation of Dependences Based on Empirical Data*. Springer New York, 1982.

[45]   R. Anand, V. Pratap, S. Kirar, and K. Burse, "K-Fold Cross Validation and Classification Accuracy of PIMA Indian Diabetes Data Set Using Higher Order Neural Network and PCA," *Int. J. Soft Comput. Eng.*, vol. 2, no. 6, pp. 436–438, 2013.

[46]   W. O. Kz, "Detection of Exudates in Digital Colour Fundus Images for the Diagnosis of Diabetic Retinopathy," Universitas Gadjah Mada, 2015.

[47]   M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J. Inf. Eng. Appl.*, vol. 3, no. 10, pp. 27–38, 2013.

[48]   V. García, R. A. Mollineda, and J. S. Sánchez, "Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions," in *Pattern Recognition and Image Analysis: 4th Iberian Conference, IbPRIA 2009 P{ó}voa de Varzim, Portugal, June 10-12, 2009 Proceedings*, H. Araujo, A. M. Mendonça, A. J. Pinho, and M. I. Torres, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 441–448.

[49]   P. S. Foundation, "Lexical Analysis." [Online]. Available: https://docs.python.org/3/reference/lexical_analysis.html#string-and-bytes-literals. [Accessed: 08-Jan-2013].

[50]   S. S. Keerthi, S. Sundararajan, K.-W. Chang, C.-J. Hsieh, and C.-J. Lin, "A sequential dual method for large scale multi-class linear svms," in *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, 2008, p. 408.