DAFTAR PUSTAKA

[1]   A. Azzalini and B. Scarpa, *Data Analysis and Data Mining an Introduction*. New York, USA: Oxford University Press, Inc., 2012.

[2]   A. O'Driscoll, J. Daugelaite, and R. D. Sleator, "'Big data', Hadoop and cloud computing in genomics," *J. Biomed. Inform.*, vol. 46, no. 5, pp. 774–781, 2013.

[3]   M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mob. Networks Appl.*, vol. 19, no. 2, pp. 171–209, 2014.

[4]   B. Lublinsky, K. T. Smith, and A. Yakubovich, *Professional Hadoop Solutions*. Indiana, CA: John Wiley & Sons, Inc., 2013.

[5]   A. Gupta, *Apache Mahout Clustering Designs*, 1st ed. Birmingham, UK: Packt Publishing Ltd, 2015.

[6]   R. M. Esteves and C. Rong, "Using Mahout for Clustering Wikipedia's Latest Articles: A Comparison between K-means and Fuzzy C-means in the Cloud," in *2011 IEEE Third International Conference on Cloud Computing Technology and Science*, 2011, pp. 565–569.

[7]   A. Huang, "Similarity measures for text document clustering," *Proc. New Zeal. Comput. Sci. Res. Student Conf. 2008*, no. April, pp. 49–56, 2008.

[8]   J. Ghosh and A. Strehl, "Similarity-Based Text Clustering: A Comparative Study," *Group. Multidimens. Data*, no. ii, pp. 73–97, 2006.

[9]   A. Strehl, E. Strehl, J. Ghosh, and R. Mooney, "Impact of Similarity Measures on Web-page Clustering," in *In Workshop on Artificial Intelligence for Web Search (AAAI 2000)*, 2000, pp. 58–64.

[10]   A. Rangrej, S. Kulkarni, and A. V. Tendulkar, "Comparative study of clustering techniques for short text documents," in *Proceedings of the 20th international conference companion on World wide web - WWW '11*, 2011, pp. 111–112.

[11]   E. Jain and S. K. Jain, "Using Mahout for clustering similar Twitter users: Performance evaluation of k-means and its comparison with fuzzy k-means," *Proc. - 5th IEEE Int. Conf. Comput. Commun. Technol. ICCCT 2014*, pp. 29–33, 2015.

[12]   E. Jain and S. K. Jain, "Categorizing Twitter users on the basis of their interests using Hadoop/Mahout platform," in *2014 9th International Conference on Industrial and Information Systems (ICIIS)*, 2014, pp. 1–5.

[13]   S. Panda, S. Sahu, P. Jena, and S. Chattopadhyay, "Comparing fuzzy-C means and K-means clustering techniques: A comprehensive study," *Adv. Intell. Soft Comput.*, vol. 166 AISC, no. VOL. 1, pp. 451–460, 2012.

[14]   S. Ghosh and S. K. S. Dubey, "Comparative analysis of k-means and fuzzy c-means algorithms," *Ijacsa*, vol. 4, no. 4, pp. 35–38, 2013.

[15]   A. Sathish and J. M. Sundaram, "A Comparative Study on K-Means and Fuzzy C-Means Algorithm for Breast Cancer Analysis," vol. 4, no. 1, pp. 54–58, 2014.

[16]   S. K. Sahu and S. K. Jena, "A Study of K-Means and C-Means Clustering Algorithms for Intrusion Detection Product Development," *Int. J. Innov. Manag. Technol.*, vol. 5, no. 3, pp. 207–213, 2014.

[17]   S. Madhukumar and N. Santhiyakumari, "Evaluation of k-Means and fuzzy C-means segmentation on MR images of brain," *Egypt. J. Radiol. Nucl. Med.*, vol. 46, no. 2, pp. 475–479, 2015.

[18]   P. Muniz De Avila *et al.*, "Comparing K-Means and Mean Shift Algorithms Performance Using Mahout in a Private Cloud Environment," *J. Commun. Comput.*, vol. 11, pp. 45–51, 2014.

[19]   L. Sahu and B. R. Mohan, "An improved K-means algorithm using modified cosine distance measure for document clustering using Mahout with Hadoop," in *2014 9th International Conference on Industrial and Information Systems (ICIIS)*, 2014, pp. 1–5.

[20]   M. a Nadaf and S. S. Patil, "Performance Evaluation of Categorizing Technical Support Requests Using Advanced K-Means Algorithm," *2015 IEEE Int. Adv. Comput. Conf.*, pp. 409–414, 2015.

[21]   S. Owen, R. Anil, T. Dunning, and E. Friedman, *Mahout in Action*, 1st ed. New York, US: Manning Publications Co., 2011.

[22]   W. Dai, "An Improved Hybrid Canopy-Fuzzy C-Means Clustering Algorithm Based on MapReduce Model," *J. Comput. Sci. Eng. Vol. 10, No. 1, March 2016, pp. 1-8 An*, vol. 10, no. 1, pp. 1–8, 2016.

[23]   S.Sathya and N.Rajendran, "A Review on Text Mining Techniques," in *International Journal of Computer Science Trends and Technology (IJCST)*, 2015, vol. 3, no. 5, pp. 274–283.

[24]   F. Xhafa, A. Bogza, S. Caballe, and L. Barolli, "Apache Mahout's k-Means vs. Fuzzy k-Means Performance Evaluation," *2016 8Th Int. Conf. Intell. Netw. Collab. Syst.*, pp. 110–116, 2016.

[25]   S. Shukla and S. Naganna, "A Review on K-means Data Clustering

Approach," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 17, pp. 1847–1860, 2014.

[26] J. Li and H. W. Lewis, "Fuzzy Clustering Algorithms — Review of the Applications," in *2016 IEEE International Conference on Smart Cloud (SmartCloud)*, 2016, pp. 282–288.

[27] A. Holmes, *Hadoop in Practics*, 1st ed. New York: Manning Publications Co., 2012.

[28] T. White, *Hadoop : The Definitive Guide*, 4th ed. Sebastopol, CA: O'Reilly Media, Inc., 2015.

[29] E. Sammer, *Hadoop Operations*. Sebastopol: O'Reilly Media, Inc., 2012.

[30] Apache, "Apache Mahout: Scalable machine learning and data mining." [Online]. Available: http://mahout.apache.org/. [Accessed: 20-Mar-2017].

[31] X. Zhang, J. Zhao, and Y. LeCun, "Character-level Convolutional Networks for Text Classification," *Comput. Res. Repos.*, vol. abs/1509.0, pp. 3057–3061, 2016.

[32] Hortonwork, "Determine HDP Memory Configuration Settings," 2017. [Online]. Available: https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.4.3/bk_installing_manually_book/content/determine-hdp-memory-config.html. [Accessed: 12-Feb-2017].