

Slow Learner Prediction Using Multi-Variate Naïve Bayes Classification Algorithm

Shiwani Rana^{*}, Roopali Garg

Department of Information Technology, UIET, Panjab University, Chandigarh, India.

Received 05 September 2016; received in revised form 21 November 2016; accepted 02 December 2016

Abstract

Machine Learning is a field of computer science that learns from data by studying algorithms and their constructions. In machine learning, for specific inputs, algorithms help to make predictions. Classification is a supervised learning approach, which maps a data item into predefined classes. For predicting slow learners in an institute, a modified Naïve Bayes algorithm implemented. The implementation is carried sing Python. It takes into account a combination of likewise multi-valued attributes. A dataset of the 60 students of BE (Information Technology) Third Semester for the subject of Digital Electronics of University Institute of Engineering and Technology (UIET), Panjab University (PU), Chandigarh, India is taken to carry out the simulations. The analysis is done by choosing most significant forty-eight attributes. The experimental results have shown that the modified Naïve Bayes model has outperformed the Naïve Bayes Classifier in accuracy but requires significant improvement in the terms of elapsed time. By using Modified Naïve Bayes approach, the accuracy is found out to be 71.66% whereas it is calculated 66.66% using existing Naïve Bayes model. Further, a comparison is drawn by using WEKA tool. Here, an accuracy of Naïve Bayes is obtained as 58.33 %.

Keywords: classification, clustering, confusion matrix, multi-variate, naïve Bayes, supervised machine learning, unsupervised machine learning, WEKA tool

References

- [1] S. Singh and S. P. Lal, "Educational courseware evaluation using machine learning techniques," Proc. IEEE Conference on e-Learning, e-Management and e-Services (IC3e 13), IEEE Press, Dec. 2013, pp. 73-78.
- [2] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," Science, vol. 349, pp. 255-260, July 2015.
- [3] M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of machine learning, London: MIT Press, 2012.
- [4] H. Bydovska and L. Popelínský, "Predicting student performance in higher education," Proc. IEEE Workshop on Database and Expert Systems Applications (DEXA 13), IEEE Press, Aug. 2013, pp. 141-145.
- [5] C. Anuradha and T. Velmurugan, "A data mining based survey on student performance evaluation system," Proc. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC 14), IEEE Press, Dec. 2014, pp. 452-456.
- [6] K. Koile, A. Rubin, S. Chapman, M. Kliman, and L. Ko, "Using machine analysis to make elementary students' mathematical thinking visible," Proc. International Conference on Learning Analytics & Knowledge (LAK 16), ACM, Apr. 2016, pp. 524-525.
- [7] B. M. McLaren, et al., "Using machine learning techniques to analyze and support mediation of student e-discussions," Frontiers in Artificial Intelligence and Applications, vol. 158, pp. 331-338, Jun. 2007.

^{*} Corresponding author. E-mail address: shiwani_rana40@gmail.com

- [8] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, pp. 131-163, Nov. 1997.
- [9] L. Jiang, H. Zhang, Z. Cai, and D. Wang, "Weighted average of one-dependence estimators," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 24, pp. 219-230, Jun. 2012.
- [10] L. Jiang, H. Zhang, and Z. Cai, "A novel Bayes model: Hidden Naive Bayes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, pp. 1361-1371, Oct. 2009.
- [11] A. B. E. D. Ahmed and I. S. Elaraby, "Data mining: a prediction for student's performance using classification method," *World Journal of Computer Application and Technology*, vol. 2, pp. 43-47, Feb. 2014.
- [12] A. Peña-Ayala, "Educational data mining: a survey and a data mining-based analysis of recent works," *Expert Systems with Applications*, vol. 41, pp. 1432-1462, Mar. 2014.
- [13] M. I. Lopez, J. M. Luna, C. Romero, and S. Ventura, "Classification via clustering for predicting final marks based on student participation in forums," *International Conference on Educational Data Mining (EDM 12)*, Jun. 2012, pp. 148-151.
- [14] M. Mayilvaganan and D. Kalpanadevi, "Comparison of classification techniques for predicting the performance of student's academic environment," *Proc. IEEE International Conference Communication and Network Technologies (ICCNT 14)*, IEEE Press, Dec. 2014, pp. 113-118.
- [15] J. Willems, "Using Learning styles data to inform learning design: a study comparing undergraduates, postgraduates and e-educators," *Australasian Journal of Educational Technology*, vol. 27, pp. 863- 880, Jan. 2011.
- [16] M. Peng, et al., "Central topic model for event-oriented topics mining in microblog stream," *Proc. International Conference on Information and Knowledge Management (CIKM 15)*, ACM, Oct. 2015, pp. 1611-1620.
- [17] J. Huang, et al., "A probabilistic method for emerging topic tracking in microblog stream," *World Wide Web Internet and Web Information Systems*, vol. 23, pp. 1-26, Apr. 2016.
- [18] S. Rana and R. Garg, "Evaluation of student's performance of an institute using clustering algorithms," *International Journal of Applied Engineering Research*, vol. 11, pp. 3605-3609, May 2016.
- [19] S. Singh and V. Kumar, "Performance analysis of engineering students for recruitment using classification data mining techniques," *International Journal of Computer Science and Engineering Technology*, vol. 3, pp. 31-37, Feb. 2013.
- [20] R. Sison and M. Shimura, "Student modeling and machine learning," *International Journal of Artificial Intelligence in Education*, vol.9, pp. 128-158, July 1998.
- [21] H. Lakkaraju, et al., "A machine learning framework to identify students at risk of adverse academic outcomes," *Proc. International Conference on Knowledge Discovery and Data Mining (KDD 15)*, ACM, Aug. 2015, pp. 1909-1918.
- [22] G. Kaur and N. Oberoi, "Naive Bayes classifier with modified smoothing techniques for better spam classification," *International Journal of Computer Science and Mobile Computing*, vol. 3, pp. 869-878, Oct. 2014.
- [23] S. Sivakumari, R. P. Priyadarsini, and P. Amudha, "Accuracy evaluation of C4.5 and Naive Bayes classifiers using attribute ranking method," *International Journal of Computational Intelligence Systems*, vol. 2, pp. 60-68, Mar. 2009.
- [24] P. Meedech, N. Iam-On, and T. Boongoen, "Prediction of student dropout using personal profile and data mining approach," *Intelligent and Evolutionary Systems*, vol. 5, pp. 143-155, May 2016.
- [25] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," *Applied Artificial Intelligence*, vol. 18, pp. 411-426, May 2004.
- [26] C. G. Nespereira, E. Elhariri, N. El-Bendary, A. F. Vilas, and R. P. D. Redondo, "Machine learning based classification approach for predicting student's performance in blended learning," *Proc. International Conference on Advanced Intelligent System and Informatics (AISI 15)*, Springer, Nov. 2015, pp. 47-56.
- [27] C. Romero, M. I. López, J. M. Luna, and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," *Computers and Education*, vol. 68, pp. 458-472, Oct. 2013.
- [28] G. I. Webb, J. R. Boughton, and Z. Wang, "Not so Naive Bayes: aggregating one-dependence estimators," *Machine learning*, vol. 58, pp. 5-24, Jan. 2005.
- [29] R. Kohavi, "Scaling up the accuracy of Naïve Bayes classifiers: a decision-tree hybrid," *Proc. International Conference on Knowledge Discovery and Data Mining (KDD 96)*, ACM, Aug. 1996, pp. 202-207.