

$$Recall = \frac{20592}{3874 + 20592} = 84.2\%$$

Obviously, the XGBoost algorithm improved the prediction accuracy from the level of the preliminary model.

5. CONCLUSIONS

This paper proposes a hybrid prediction model for customer churn based on logistics regression and XGBoost algorithm model. More than twenty indices were selected from the dimensions like order information and customer profile as the independent variables for the hybrid model. The model was applied to predict the churn state of customers of an actual e-commerce platform. Judging by accuracy, precision and recall, it can be seen that the hybrid model can predict customer churn more accurately than logistic regression.

REFERENCES

- [1] Chen, D., Sain, S.L., Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(3): 197-208. <https://doi.org/10.1057/dbm.2012.17>
- [2] Chang, H., Jamin, S., Willinger, W. (2001). Inferring AS-level Internet topology from router-level path traces. *International Society for Optics and Photonics, Scalability and traffic control in IP networks*, 4526: 196-207. <https://doi.org/10.1117/12.434395>
- [3] Lee, N., Kim, J.M. (2010). Conversion of categorical variables into numerical variables via Bayesian network classifiers for binary classifications. *Computational Statistics & Data Analysis*, 54(5): 1247-1265. <https://doi.org/10.1016/j.csda.2009.11.003>
- [4] Shaker, A., Senge, R., Hüllermeier, E. (2013). Evolving fuzzy pattern trees for binary classification on data streams. *Information Sciences*, 220: 34-45. <https://doi.org/10.1016/j.ins.2012.02.034>
- [5] Unler, A., Murat, A. (2010). A discrete particle swarm optimization method for feature selection in binary classification problems. *European Journal of Operational Research*, 206(3): 528-539. <https://doi.org/10.1016/j.ejor.2010.02.032>
- [6] Ebenuwa, S.H., Sharif, M.S., Alazab, M., Al-Nemrat, A. (2019). Variance ranking attributes selection techniques for binary classification problem in imbalance data. *IEEE Access*, 7, 24649-24666. <https://doi.org/10.1109/ACCESS.2019.2899578>
- [7] Shao, Y.H., Chen, W.J., Deng, N.Y. (2014). Nonparallel hyperplane support vector machine for binary classification problems. *Information Sciences*, 263: 22-35. <https://doi.org/10.1016/j.ins.2013.11.003>
- [8] Xu, Z., Watada, J., Wu, M., Ibrahim, Z., Khalid, M. (2014). Solving the imbalanced data classification problem with the particle swarm optimization based support vector machine. *IEEJ Transactions on Electronics, Information and Systems*, 134(6): 788-795. <https://doi.org/10.1541/ieejieiss.134.788>
- [9] Huang, Y., Wang, Y. (2012). Decision tree classification based on naive Bayesian and ID3 algorithm. *Computer Engineering*, 38(14): 41-43.
- [10] Ahmed, M., Afzal, H., Majeed, A., Khan, B. (2017). A survey of evolution in predictive models and impacting factors in customer churn. *Advances in Data Science and Adaptive Analysis*, 9(3): 1750007. <https://doi.org/10.1142/S2424922X17500073>
- [11] Ju, C.H., Lu, Q.B., Guo, F.P. (2013). E-commerce customer churn prediction model combined with individual activity. *Systems Engineering-Theory & Practice*, 33(1): 141-150.
- [12] Chan, K.K., Misra, S. (1990). Characteristics of the opinion leader: A new dimension. *Journal of Advertising*, 19(3): 53-60. <https://doi.org/10.1080/00913367.1990.10673192>
- [13] Carver, T., Harris, S. R., Berriman, M., Parkhill, J., McQuillan, J.A. (2011). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, 28(4): 464-469. <https://doi.org/10.1093/bioinformatics/btr703>
- [14] Fathian, M., Hoseinpoor, Y., Minaei-Bidgoli, B. (2016). Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods. *Kybernetes*, 45(5): 732-743. <https://doi.org/10.1108/K-07-2015-0172>
- [15] Sharma, A., Panigrahi, P.K. (2013) A neural network based approach for predicting customer churn in cellular network services. *International Journal of Computer Applications*, 27(11): 26-31. <https://doi.org/10.5120/3344-4605>
- [16] Keramati, A., Azadeh, A., Mohammadi, M., Rostami, H. (2011). Identification of customer churn determinants using censored log file data in the Iranian mobile telecommunications service industry. *International Journal of Electronic Customer Relationship Management*, 5(2): 111-129. <https://doi.org/10.1504/IJECRM.2011.041261>
- [17] Su, Q., Shao, P., Ye, Q. (2012). The analysis on the determinants of mobile VIP customer churn: A logistic regression approach. *International Journal of Services Technology and Management*, 18(1-2): 61-74. <https://doi.org/10.1504/IJSTM.2012.049016>