Class imbalance is a major problem that is often experienced when dealing with rare events, such as churn recognition in the mobile telecommunications industry. In this work, various strategies of addressing the problem are studied and a demonstration of how under-sampling and Synthetic Minority Oversampling Technique (SMOTE) can be used to address the problem is given. The two techniques are implemented individually first, and then we take the hybrid approach by combining both SMOTE and undersampling. For performance evaluation, two predictive techniques, C4.5 decision tree and Naïve Bayes classifier with 10-fold cross validation are used. TPR and FPR values are obtained and used to generate ROC curves from which AUC values are calculated and performance comparison of the three techniques is performed. Results show that the hybrid approach achieves better performance.

References

Handling Class Imbalance in Mobile Telecoms Customer Churn Prediction


- P. N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Pearson Education Asia Inc., 2006.


- J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, 2012.


Index Terms

Computer Science
Algorithms
Keywords
Class Imbalance  Customer Churn  Over-sampling  Under-sampling  Prediction