Abstract

Customer churn is a common problem encountered in the telecommunications industry. Customer churn is defined as the tendency of the customer to stop doing business with a company. But there are very few churn customers available. Lack of data that indicates that the customer including churn customer causes an imbalanced data problem. In this final project the author makes a system that can perform related data imbalanced handling using SMOTE (Synthetic Minority Over-sampling Technique) method. Classifier used to determine a customer whether including churn or not, using Improved KNN Algorithm based on Kernel Method. This method is an improvement of the KNN Standard method. Where in the KNN Standard method the classification process is done by looking at a number of nearest neighbes k, and will be classified by the largest number of classes on a number of its nearest neighbes k. The classifier is tested using 3 Kernel functions and 40 parameter combinations to find the highest performance. The highest performance obtained from the combination of parameters is measured using f1-measure and accuracy sequentially in the data without smote, smote 1: 3, smote 1: 2, smote 3: 4, and smote 1: 1, ie: 0.314 & 97.58% 0.449 & 94.55%, 0.413 & 93.70%, 0.382 & 92.74% and 0.363 & 92.08%.

Keywords: Churn Prediction, Over-sampling, SMOTE (Synthetic Minority Over-sampling Technique), Improved KNN Algorithm Based on Kernel Method.