عنوان مقاله:

Extracting Rules from Imbalanced Data: The Case of Credit Scoring

نویسندگان:

Seyed Mahdi Sadatrasoul - *Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran*

Mohammad Reza Gholamian - *Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran*

Kamran Shahanaghi - *Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran*

خلاصه مقاله:

Credit scoring is an important topic, and banks collect different data from their loan applicant to make an appropriate and correct decision. Rule bases are of more attention in credit decision making because of their ability to explicitly distinguish between good and bad applicants. The credit scoring datasets are usually imbalanced. This is mainly because the number of good applicants in a portfolio of loan is usually much higher than the number of loans that default. This paper use previous applied rule bases in credit scoring, including RIPPER, OneR, Decision table, PART and C4.5 to study the reliability and results of sampling on its own dataset. A real database of one of an Iranian export development bank is used and, imbalanced data issues are investigated by randomly Oversampling the minority class of defaulters, and three times under sampling of majority of non-defaulters class. The performance criterion chosen to measure the reliability of rule extractors is the area under the receiver operating characteristic curve (AUC), accuracy and number of rules. Friedman's statistic is used to test for significance differences between techniques and datasets. The results from study show that PART is better and good and bad samples of data affect its results less

کلمات کلیدی:

Credit Scoring; Banking Industry; Rule Extraction; Imbalanced Data; Sampling

لینک ثابت مقاله در پایگاه سیویلیکا:

https://civilica.com/doc/571097