



Título: Use of random forest with cost-sensitive dynamic selection

techniques to imbalanced credit scoring datasets

Data: 19/06/2019

Horário: 08:30h

Local: Sala de Seminários - Bloco 952

Resumo:

The profitability of the banks highly depends on the models used to decide on the customer's loans. State of the art credit scoring models is based on machine learning methods. These methods need to cope with the problem of imbalanced classes since credits coring datasets usually contain many paid loans, and few not paid ones (defaults). Recently, dynamic selection approaches combined with preprocessing techniques have been evaluated for imbalanced datasets. However, previous works only evaluate pre-processing techniques combined with pool generator ensembles. For this reason, we propose to

combine random forest and random undersampling with a cost-sensitive version of dynamic selection techniques. We assess the prediction performance by using seven real-world credit scoring datasets with different levels of imbalanced ratio. Experimental results show that cost-sensitive dynamic selection combined with random forest and random undersampling improves the classification performance concerning the static ensembles and the other dynamic selection combinations.

Banca:

- Prof. Dr. José Antônio Fernandes de Macedo (MDCC/UFC - Orientador)
- Prof. Dr. César Lincoln Cavalcante Mattos (MDCC/UFC)
- Prof. Dr. Chiara Renso (ISTI - CNR)
- Prof. Dr. Franco Maria Nardini (ISTI - CNR)