

Random Forest

pour prédire les faillites d'entreprises

1 Présentation du problème

Le Service d'information sur les marchés émergents (Emerging Markets Information Service ou EMIS) possède une base de données contenant des informations sur les marchés émergents du monde entier. A partir de cette base, des données concernant la faillite d'entreprises polonaises ont été extraites [1]. Les sociétés en faillite ont été analysées au cours de la période 2000-2012, tandis que les sociétés encore en activité ont été évaluées de 2007 à 2013. Cinq cas de classification ont été distingués, en fonction de la période de prévision.

2 Travail demandé

Le but du projet est de concevoir et d'entraîner une forêt aléatoire constituée d'arbres de décision pour prédire si une entreprise fera faillite. On s'intéressera aux données de la 5^e année uniquement.

2.1 Réalisation

Les données sont librement accessibles sur le UCI Machine Learning Repository¹. Vous y trouverez plus de détails sur le problème et les données.

Le choix du logiciel, de la bibliothèque ou du langage de programmation est libre.

2.2 Rapport

Un court rapport devra exposer la méthode et les paramètres que vous avez choisis, vos résultats et les améliorations possibles. Le rapport et les sources (lesquelles peuvent correspondre à la chaîne de traitement dans knime par exemple) doivent être déposés sur le site arche dédié au module Statistique Avancée au plus tard le vendredi 8 mars à 18h. Merci de nommer votre archive avec votre nom sur le format : Faillites-Nom.zip

Références

[1] Zieba, M., Tomczak, S. K., Tomczak, J. M. (2016). Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction. Expert Systems with Applications

1. <https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data>