

Usage non classificatoire d'arbres de classification : enseignements d'une analyse de la participation féminine à l'emploi en Suisse

Fabio B. Losa*, Pau Origoni*, Gilbert Ritschard**

*Office statistique du canton du Tessin, Bellinzona, Suisse

fabio.losa@ti.ch ; pau.origoni@ti.ch

**Département d'économétrie, Université de Genève

gilbert.ritschard@themes.unige.ch

Résumé. Cet article présente une application en grandeur réelle des arbres de classification dans un contexte non classificatoire. Les arbres générés visent à mettre en lumière les différences régionales dans la façon dont les femmes décident de leur participation au marché du travail. L'accent est donc mis sur la capacité descriptive plutôt que prédictive des arbres. L'application porte sur des données relatives à la participation féminine au marché du travail issues du Recensement Suisse de la Population de l'an 2000. Ce vaste ensemble de données a été analysé en deux phases. Un premier arbre exploratoire a mis en évidence la nécessité de procéder à des études séparées pour les non mères, les mères mariées ou veuves, et les mères célibataires ou divorcées. Nous nous limitons ici aux résultats de ce dernier groupe, pour lequel nous avons généré un arbre séparé pour chacune des trois régions linguistiques principales. Les arbres obtenus font apparaître des différences culturelles fondamentales entre régions. Du point de vue méthodologique, la principale difficulté de cet usage non classificatoire des arbres concerne leur validation, puisque le taux d'erreur de classification généralement retenu perd tout son sens dans ce contexte. Nous commentons cet aspect et illustrons l'usage d'alternatives plus pertinentes et facilement calculables.

1 Introduction

Les arbres de décision sont, depuis leur popularisation par Breiman et al. (1984), devenus des outils multivariés privilégiés pour prédire la valeur de variables continues ou la classe de variables catégorielles à partir d'un ensemble de prédicteurs. On parle d'*arbre de régression* quand l'attribut à prédire est quantitatif et d'*arbre de classification* lorsqu'il est catégoriel. Bien que leur but premier soit la prédiction et la classification, les arbres présentent bien d'autres intérêts, comme méthode exploratoire pour partitionner et identifier des structures locales dans les bases de données, mais aussi comme alternative aux méthodes statistiques classiques comme la régression linéaire ou logistique par exemple (Wilkinson, 1992).

Cette contribution illustre cet usage *non-classificatoire* des arbres de classification en présentant une application réelle sur des données relatives à la participation féminine