

Krzysztof Karpio, Piotr Łukasiewicz, Arkadiusz Orłowski, Tomasz Ząbkowski

Discovering and analyzing income determinants using decision trees

A problem of selecting some microeconomic factors which could influence income distributions is investigated. We concentrate on potential income determinants for Polish households using data from 2000 to 2010. Decision trees, based on the random subsets of data and empowered by the entropy measure, are constructed and interpreted. Within the applied procedure the following steps are performed for each investigated year: selection of determinants, evaluation of strength of influence and its error for each determinant, ordering of selected determinants according to strength of influence. Possible relationships of this approach with the maximum entropy methods of statistical physics are pointed out and discussed. Results obtained for all investigated years are presented and thoroughly compared.