Abstract

The fundamental population data are needed for every country for purposes of planning, development, and improvement. Census data can provide the basic population data of any country. Moreover, they are rich with lots of hidden information that can be used for machine learning and data mining tasks in order to provide services for country's social and economic development. This paper is focused on the applications of data mining and machine learning in census data to classify the annual income. It aims to show a systematic comparison to examine and evaluate three supervised learning classifiers. The classifiers that have been targeted are decision trees, random forests, and artificial neural networks. The main aims are to explore not only the classifiers properties and the impact of the attributes on the evaluation, but also, evaluate their classification performance under certain conditions to understand how the performance of the models changes over different experiments which potentially provide a guidance to help researchers to determine the most suitable classifier in census data.

References

Index Terms

Computer Science

Algorithms
Keywords

Census Data, Data Mining, Classification, Supervised Learning, Decision Trees, Random Forests, Artificial Neural Networks, Performance Metrics.