Big Data Analytics Project : Dataset Adult from UCI

Mohammed Meftah, Raunaq Paul and Binxiang Xiang

16th January 2017

Contents

1   Introduction

For this project, we have decided to work on the dataset Adult from UCI. This is a classification dataset with two classes, as we have to understand and predict whether a person earns less or more than 50k dollars a year. Section 2 describes in details the used dataset Adult. Section 3 highlights how we first processed data. Section 4 shows some statistical analysis on the dataset. Section 5 presents the methodology and the performances of different algorithms (list of tried algo to add) on the prediction of our target variable.

2   Dataset description

3   Data processing

4   Analysis

5   Methodology and results

6   Conclusion