

List of team members: Henry Qiu, Goktug Cengiz

Our proposal for the project is to carry out a problem that consists in determining whether a person can make over 50.000 dollars in a year, taking into account some known information like its social situations.

There are many factors that can determine the economic situation of a person, many times we make questions as: How are the people who earns a lot of money? How should I be to be able to earn a lot of money? Solving this problem will ask these kind of questions. Analyzing this data set can also contribute valuable information for sociologists to study the factors that affects the success of a person.

Our data set is called **Adult Data Set** got from the *UCI Machine Learning Repository*[1] . This data set contains a large amount of attributes, some of them categorical and some of them numerical, this will make the task more interesting. The amount of instances is also significant, with big ammount of instances we will have rich content of information so that the learning part can be much more complex. What is more, there are also some missing values among the data so that it will be perfect for us to do the data preprocessing part.

We chose this data set because this problem has awakened in us huge interests in analyzing how can social factors impacts people's life. We found this data set really easy to understand, in comparison with other areas, the concepts involved in social areas are much easier to interpret, they do not require deep knowledge of vocabulary as other areas, such as biology, chemistry, and so on. What is more, this data set is very appropriate for our studies for the reasons we explained above: the amount of instances, the amount and features of the variables, the amount of missing values are all very suitable for our learning.

## References

<https://archive.ics.uci.edu/ml/datasets/Adult> - Adult Data Set - By Ronny Kohavi and Barry Becker - 1996-05-01.