# Задание: построить классификатор для заданного набора данных и проверить качество его работы на тестовом множестве

Источник и описание данных: <http://archive.ics.uci.edu/ml/datasets/Adult>

На основании собранных данных необходимо определить, сможет ли человек заработать более 50000 в год.

Описание переменных и их свойства:

|  |  |  |
| --- | --- | --- |
| **Имя переменной** | **Тип переменной** | **Значения** |
| age | непрерывная |  |
| workclass | категориальная | Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked. |
| fnlwgt | непрерывная |  |
| education | категориальная | Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool. |
| education-num | непрерывная |  |
| marital-status | категориальная | Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse. |
| occupation | категориальная | Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces. |
| relationship | категориальная | Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried. |
| race | категориальная | White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black. |
| sex | категориальная | Female, Male. |
| capital-gain | непрерывная |  |
| capital-loss | непрерывная |  |
| hours-per-week | непрерывная |  |
| native-country | категориальная | United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands. |

**Для решения задачи использовать деревья классификации и метод ближайшего соседа**.