

Buổi 4: Bài tập

Trần Văn Lộc

`ml.course.202k@gmail.com`

`https://github.com/ml-course-202k`

Ngày 5 tháng 5 năm 2020

Với data titanic.csv hãy chỉ định rõ kiểu dữ liệu và giải thích tại sao, chuẩn hóa dữ liệu

Mô tả dữ liệu:

PassengerId: Id của hành khách

Survived: Có được cứu sống hay không, 1 -> có, 0 -> không

PClass: Địa vị xã hội Name: Tên của hành khách

Sex: Giới tính của hành khách

Age: Tuổi của hành khách

Ticket: Số vé

Fare: Giá vé

Embarked: Địa điểm khởi hành, C = Cherbourg, Q = Queenstown, S = Southampton

Với data bank.csv hãy chỉ định rõ kiểu dữ liệu và giải thích tại sao, chuẩn hóa dữ liệu

Mô tả dữ liệu:

age: Tuổi khách hàng

job: Công việc

marital: Tình trạng hôn nhân

education: Giáo dục

balance: Ngân sách (đơn vị: đô la)

housing: Có nhà hay không

loan: Có khoản vay nào hay không?

contact: hình thức liên lạc

day: Ngày liên lạc

month: Tháng liên lạc