

تمرین سری سوم هوش مصنوعی

پدرام طاهری، ۶۱۰۳۹۵۱۲۳

پیش‌پردازش:

ستون‌های "Capital loss", "Capital gain", "Education" حذف شدند. تحصیلات به دلیل تکراری بودن و دوتای دیگر به دلیل تعداد زیاد داده‌های گم‌شده و ناموجود. سپس داده‌های رشته‌ای به دسته‌بندی و سپس به اعداد تبدیل شدند (هر عدد نشان‌دهنده یک کلاس).

تقسیم‌بندی داده یادگیری و تست به صورت ۶۶ به ۳۳ بود.

کلاس‌بندی:

با استفاده از GuassianNB و با استفاده از کتابخانه scikit انجام شد. امتیازها به شکل زیر است:

```
X_train, X_test, y_train, y_test = load_and_clean_data()  
train_and_report(X_train, X_test, y_train, y_test)
```

Training score: 0.7918404767361907

Test score: 0.7909919970221477