

## PRÁCTICA ADULT

Tenéis que crear el mejor modelo posible que prediga el *income* de una serie de usuarios.

Esta es la base de datos (que también tenéis en el campus)

<https://archive.ics.uci.edu/ml/datasets/adult>

Esta son las headers:

```
headers = ['age', 'workclass', 'fnlwgt', 'education', 'education-num', 'marital-status', 'occupation', 'relationship', 'race', 'sex', 'capital-gain', 'capital-loss', 'hours-per-week', 'native-country', 'income']
```

Usar cross\_validation con cv=5 como en el siguiente ejemplo.

```
from sklearn.cross_validation import RandomForestClassifier
from sklearn.cross_validation import cross_val_score

X = df[['age', 'education-num', 'capital-gain']]
y = df.income_

forest = RandomForestClassifier()
scores = cross_val_score(forest, X, y, cv=5)
print("Accuracy: %.2f (+/- %.2f)" % (scores.mean(), scores.std() * 2))

Accuracy: 0.8681 (+/- 0.01)
```

Tenéis que seleccionar diferentes algoritmos, tunear los hiperparámetros, modificar features hasta que conseguir el mejor modelo. Como referencia yo consigo:

Accuracy: 0.8681 (+/- 0.01)