

تمرین ششم درس داده کاوی

- به زبان برنامه نویسی پایتون برنامه ای بنویسید که موارد خواسته شده زیر را انجام دهد.
- موارد خواسته شده را به دقت انجام دهید.
- از خروجی های بدست آمده در هر مرحله عکس بگیرید و در فایل پاسخ قرار دهید. پاسخ های خود را به طور کامل به صورت صوتی شرح دهید.
- فایل نهایی توضیحات به همراه کدها و خروجی های بدست آمده را به صورت یک فایل فشرده (rar یا zip) در لینک زیر آپلود کنید. اگر ویدیو قرار داده می شود بایستی حجم آن تا حد امکان کاهش داده شود.
<https://cloud.vru.ac.ir/s/3y4PpkqgiYJkmC6>
- به کسی که بهترین و کاملترین محتوا برای هر تمرین را تولید کند (هم از نظر صحت پاسخ و هم از نظر گرافیکی و هم ارزش آموزشی برای دیگران) نمره اضافه نیز در نظر گرفته خواهد شد.

درخت تصمیم

- دادگان penguins.csv که به همراه تمرین ارسال شده است را لود کرده و سپس پردازش های خواسته شده را انجام دهید.
- ۱- نمونه داده هایی که دارای مقادیر گمشده (missing value) هستند را از داده ها حذف کنید.
 - ۲- داده ها را به دو مجموعه train و test تقسیم کنید به طوری که ۲۰٪ داده ها در مجموعه تست باشند.
 - ۳- با استفاده از روش درخت تصمیم طبقه بندی (تشخیص گونه: Adelie, Chinstrap, Gentoo) را انجام دهید.

```
from sklearn.tree import DecisionTreeClassifier  
dtree = DecisionTreeClassifier()
```
 - ۴- کانفیوژن ماتریس و پارامترهای ارزیابی را محاسبه کنید
 - ۵- درخت تصمیم نهایی بدست آمده را نمایش دهید:
- راهنمای بیشتر: از لینک زیر کمک بگیرید.

https://www.w3schools.com/python/python_ml_decision_tree.asp

SVM

- دادگان bill_authentication.csv که به همراه تمرین ارسال شده است را لود کرده و سپس پردازش های خواسته شده را انجام دهید.
- ۱- داده ها را به دو مجموعه train و test تقسیم کنید به طوری که ۲۰٪ داده ها در مجموعه تست باشند.
 - ۲- با استفاده از روش SVM خطی، طبقه بندی (تشخیص class) را انجام دهید.

```
from sklearn.svm import SVC  
svclassifier = SVC(kernel='linear')
```
 - ۳- کانفیوژن ماتریس و پارامترهای ارزیابی را محاسبه کنید
 - ۴- با استفاده از روش SVM غیرخطی (کرنل گاوسی)، طبقه بندی (تشخیص class) را انجام دهید

```
svclassifier = SVC(kernel='rbf')
```
 - ۵- کانفیوژن ماتریس و پارامترهای ارزیابی را محاسبه کنید

شبکه های عصبی مصنوعی

دادگان bill_authentication.csv که به همراه تمرین ارسال شده است را لود کرده و سپس پردازش های خواسته شده را انجام دهید.

۱- داده ها را به دو مجموعه train و test تقسیم کنید به طوری که ۲۰٪ داده ها در مجموعه تست باشند.

۲- با استفاده از یک شبکه عصبی MLP، طبقه بندی (تشخیص class) را انجام دهید.

```
from sklearn.neural_network import MLPClassifier
```

۳- کانفیوژن ماتریس و پارامترهای ارزیابی را محاسبه کنید

۴- تعداد نرون های لایه مخفی شبکه را از تعداد ۲ تا ۲۰ تغییر دهید و برای هر مدل دقت را محاسبه و نمودار تغییرات دقت نسبت به تغییرات تعداد نرونهای لایه مخفی را رسم کنید.

موفق باشید