

تمرین پنجم 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/kZigt4aogg6aK6C>
- 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.

KNN

- 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- ویژگی های (صفت های) مختلف در داده را نرمال سازی نمایید.
 - ۴- با استفاده از روش KNN طبقه را انجام دهید. (تعداد همسایه ها یعنی k را ۵ در نظر بگیرید.
- ```
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=5)
```
- ۵- کانفیوژن ماتریس و پارامترهای ارزیابی را محاسبه کنید
  - ۶- طبقه بندی را با مقادیر مختلف k از ۱ تا ۴۰ انجام دهید و خطای بازشناسی بدست آمده را در یک نمودار نشان دهید و تفسیر کنید.
- راهنمای بیشتر: از لینک زیر کمک بگیرید مشابه فرایندی که برای دادگان iris انجام شده برای داده خود انجام دهید.
- <https://stackabuse.com/k-nearest-neighbors-algorithm-in-python-and-scikit-learn/>

## Naïve Bayes

با استفاده از داده های داده شده در جدول زیر و با استفاده از روش بیز ساده جواب سوالات زیر را بدست آورید.

|                             |      |
|-----------------------------|------|
| D1: "send us your password" | Spam |
| D2: "send us your review"   | Spam |
| D3: "review your password"  | Ham  |
| D4: "review us"             | Ham  |
| D5: "send us password"      | Spam |
| D6: "send us your account " | Spam |

\* تشریح مراحل و محاسبات الزامی است.

۱- "send account" به کدام کلاس تعلق دارد؟

۲- "review account" به کدام کلاس تعلق دارد؟

موفق باشید