

تمرین سری چهار هوش مصنوعی

پدرام طاهری، ۶۱۰۳۹۵۱۲۳

برای نرمال‌سازی، توکنایز و لماتایز کردن متن از کتابخانه اسپیسی استفاده شد. مدل‌های نیز از سایکیت.
(بالای متن‌های حذف شد (به جز فیلد ساجکت). خطوطی که با < شروع می‌شدند نیز حذف شدند و در انتها ایمیل‌ها، اعداد،
استاپ‌وردها و کاراکترهای غیر حروف نیز از متن خارج شدند. کلمات استرپ شدند و فاصله‌ها نیز از بین رفتند تا
تی‌اف‌آی‌دی‌اف به حداکثر بازدهی برسد)
نتایج در عکس زیر پیوست شده.

```
In [40]: km_model = KMeans(n_clusters=3)
km_model.fit_transform(X)

clustering = collections.defaultdict(list)
for idx, label in enumerate(km_model.labels_):
    clustering[label].append(idx)

print("Micro F-1 Score: %0.3f" % metrics.f1_score(labels, km_model.labels_, average='micro'))
print("Macro F-1 Score: %0.3f" % metrics.f1_score(labels, km_model.labels_, average='macro'))
print("Adjusted Rand-Index: %0.3f" % metrics.adjusted_rand_score(labels, km_model.labels_))

Micro F-1 Score: 0.334
Macro F-1 Score: 0.291
Adjusted Rand-Index: 0.699
```

```
In [41]: gmm_model = GaussianMixture(n_components=3, covariance_type='diag')
gmm_model.fit(X.toarray())
pred = gmm_model.predict(X.toarray())

print("Micro F-1 Score: %0.3f" % metrics.f1_score(labels, pred, average='micro'))
print("Macro F-1 Score: %0.3f" % metrics.f1_score(labels, pred, average='macro'))
print("Adjusted Rand-Index: %0.3f" % metrics.adjusted_rand_score(labels, pred))

Micro F-1 Score: 0.330
Macro F-1 Score: 0.294
Adjusted Rand-Index: 0.733
```