

# گزارش تمرین چهارم

مقایسه تیتر اخبار صفحه اول روزنامه‌ها در دو  
دهه ۷۰ و ۸۰ به کمک mallet

زهرا صادقی عدل  
۹۴۵۲۳۱۸۹

الف) تبدیل فایل داده به فرمت مطلوب :mallet  
برای این مرحله به وسیله کد پایتون موجود در گیت تیتر به تیر مطالب را جدا کرده و  
برای هر یک دو فیچر کلمه اول و کلمه اخر را مشخص نمودیم.

ب) به کمک دستورات ساده mallet که مطابق زیر میباشند الگوریتم‌های مربوطه را  
اعمال کردیم:

```
bin/mallet import-file --input /Users/zahra/desktop/mallet_text.txt --  
output labeled.mallet  
  
bin/mallet train-classifier --input labeled.mallet --training-portion  
0.9 --trainer MaxEnt --trainer NaiveBayes
```

نتیجه به این صورت است:

NaiveBayesTrainer	MaxEntTrainer,gaussianPriorVariance=1.0
Summary. train accuracy mean = 0.9233976029181865	Summary. train accuracy mean = 0.9411151641479938
Summary. test accuracy mean = 0.812206572769953	Summary. test accuracy mean = 0.8028169014084507
Summary. test precision(1385) mean = 0.8444444444444444	Summary. test precision(1385) mean = 0.8717948717948718
Summary. test precision(1375) mean = 0.803571428571428	Summary. test precision(1375) mean = 0.7873563218390804
Summary. test recall(1385) mean = 0.5352112676056338	Summary. test recall(1385) mean = 0.4788732394366197
Summary. test recall(1375) mean = 0.9507042253521126	Summary. test recall(1375) mean = 0.9647887323943662
Summary. test f1(1385) mean = 0.6551724137931035	Summary. test f1(1385) mean = 0.6181818181818182
Summary. test f1(1375) mean = 0.8709677419354838	Summary. test f1(1375) mean = 0.8670886075949367

با مقایسه نتایج مشخص میشود که تقریبا درستی و نتیجه هردو مشابه یکدیگر  
میباشد.