

بسم الله الرحمن الرحيم

فاز سوم تمرین دوم درس NLP

برای این فاز ابتدا چند فیچر را در نظر گرفتیم. این فیچرها شامل کلمه اول مصرع اول و دوم بیت اول هر نمونه، بیست کلمه به دست آمده از وردمپ فاز اول پروژه و نقش کلمه اول و آخر هر نمونه هستند.

سپس با فرمت زیر این اطلاعات را در فایل data.txt می نویسیم:

```
InstanceName label Feature1Name Feature1Value Feature2Name  
Feature2Value ...
```

برای مثال یک خط این فایل به شکل زیر می باشد :

```
14 hafez f1 نصیحت f2 جوانان f3 0 f4 0 f5 1 f6 0 f7 0 f8 0 f9 0 f10 0 f11  
0 f12 0 f13 0 f14 0 f15 0 f16 0 f17 0 f18 1 f19 0 f20 0 f21 1 f22 0  
f23 Ne f23 N
```

اعداد صفر و یک نشان دهنده وجود و یا عدم وجود واژه در نظر گرفته در نمونه می باشد.

پس از ساختن این فایل در cmd با زدن دستور زیر فایلی با فرمت `mallet` می سازیم :

```
bin\mallet train-classifier --input data.mallet
```

و بعد از آن با زدن دستور زیر نتایج `precision` و `recall` و `f1` را برای دو الگوریتم

`NaiveBayes` و `MaxEnt` به دست می آوریم :

```
bin\mallet train-classifier --input data.mallet --training-portion 0.9  
--trainer MaxEnt --trainer NaiveBayes
```

نتایج به دست آمده بدون در نظر گرفتن دو فیچر نقش کلمات اول آخر به شکل زیر می باشد :

```

Trial 0 Trainer NaiveBayesTrainer test data precision(hafez) = 0.5806451612903226
Trial 0 Trainer NaiveBayesTrainer test data precision(molana) = 0.6129032258064516
Trial 0 Trainer NaiveBayesTrainer test data recall(hafez) = 0.75
Trial 0 Trainer NaiveBayesTrainer test data recall(molana) = 0.4222222222222222
Trial 0 Trainer NaiveBayesTrainer test data F1(hafez) = 0.6545454545454547
Trial 0 Trainer NaiveBayesTrainer test data F1(molana) = 0.5
Trial 0 Trainer NaiveBayesTrainer test data accuracy = 0.5913978494623656

MaxEntTrainer,gaussianPriorVariance=1.0
Summary. train accuracy mean = 0.818362480127186 stddev = 0.0 stderr = 0.0
Summary. test accuracy mean = 0.5985663082437276 stddev = 0.0 stderr = 0.0
Summary. test precision(hafez) mean = 0.7285714285714285 stddev = 0.0 stderr = 0.0
Summary. test precision(molana) mean = 0.5550239234449761 stddev = 0.0 stderr = 0.0
Summary. test recall(hafez) mean = 0.3541666666666667 stddev = 0.0 stderr = 0.0
Summary. test recall(molana) mean = 0.8592592592592593 stddev = 0.0 stderr = 0.0
Summary. test f1(hafez) mean = 0.4766355140186917 stddev = 0.0 stderr = 0.0
Summary. test f1(molana) mean = 0.6744186046511628 stddev = 0.0 stderr = 0.0

NaiveBayesTrainer
Summary. train accuracy mean = 0.777027027027027 stddev = 0.0 stderr = 0.0
Summary. test accuracy mean = 0.5913978494623656 stddev = 0.0 stderr = 0.0
Summary. test precision(hafez) mean = 0.5806451612903226 stddev = 0.0 stderr = 0.0
Summary. test precision(molana) mean = 0.6129032258064516 stddev = 0.0 stderr = 0.0
Summary. test recall(hafez) mean = 0.75 stddev = 0.0 stderr = 0.0
Summary. test recall(molana) mean = 0.4222222222222222 stddev = 0.0 stderr = 0.0
Summary. test f1(hafez) mean = 0.6545454545454547 stddev = 0.0 stderr = 0.0
Summary. test f1(molana) mean = 0.5 stddev = 0.0 stderr = 0.0

```

	Precision (hafez)	Precision (molana)	Recall (hafez)	Recal (molana)	F1 (hafez)	F1 (molana)
Naïve Bayes	0.58	0.61	0.75	0.42	0.65	0.5
MaxEnt	0.72	0.55	0.35	0.85	0.47	0.67