

گزارش تمرین سوم :

با استفاده از feature هایی که در تمرین دوم به دست آورده بودیم از آن ها نیز استفاده می کنیم و در دایرکتوری mallet در کنار فایل mallet.py قرار می دهیم. حال طبق همیشه داده ها را از فایل ها استخراج می کنیم و آن دسته از کلماتی را که فیلتر کردیم و همچنین جزو ویژگی ها (feature ها) بودند را انتخاب می کنیم و به ازای هر فایل تمامی این کلمات با فیلتر های گفته شده را در کنار یکدیگر قرار می دهیم و به صورت یک خط در با شمارنده که استفاده کردیم در فایل mallet.txt ذخیره سازی می کنیم. نحوه ی ذخیره سازی به صورت زیر است :

Counter className features

که در آن counter همان شمارنده ای است که استفاده کردیم و className نام کلاس هایمان (رپ و پاپ) و features همان کلماتی است که از هر کلاس خارج کردیم و با features.txt تطابق داده ایم. سپس با استفاده از خود ابزار Mallet فایل mallet.txt را به عنوان ورودی به آن می دهیم و out.mallet را به عنوان خروجی از آن دریافت می کنیم.

```
(testenv)
# slytanix @ Shayans-MacBook-Pro in ~/Desktop/Courses/NLP/Projects/wordmap/Mallet on git:master x [16:33:06]
$ ../../mallet-2.0.6/bin/mallet import-file --input mallet.txt --output out.mallet
(testenv)
# slytanix @ Shayans-MacBook-Pro in ~/Desktop/Courses/NLP/Projects/wordmap/Mallet on git:master x [16:33:12]
$ ../../mallet-2.0.6/bin/mallet train-classifier --input out.mallet --training-portion 0.9 --trainer MaxEnt --trainer NaiveBayes
Training portion = 0.9
Unlabeled training sub-portion = 0.0
Validation portion = 0.0
Testing portion = 0.09999999999999998
```

در فایل mallet.txt ورودی ها به صورت متون هستند و در صورت تغییر ورودی به جملات می تواند نتایج بهتری را حاصل کند.