

در ادامه‌ی همان کد کلسیفایر naive bayes کد تولید کننده دیتا برای mallet را اضافه کردیم (در تابع mallet). ابتدا کلماتی که بیشترین تاثیر را در کلسیفایر naive bayes داشتن را پیدا کردیم و ۱/۱۰ از آن‌ها را به عنوان feature در دیتای mallet اضافه کردیم. سپس دو feature دیگر نیز اضافه کردیم که یکی طول جمله و دیگری وجود عدد در جمله بود.

با استفاده از کد زیر در ترمینال از max entropy استفاده کردیم (نام فایل ورودی mallet.txt است):

```
bin/mallet import-file --input mallet.txt --output web.out
```

```
bin/mallet train-classifier --input web.mallet --training-portion 0.9  
--trainer MaxEnt --report test:raw test:confusion train:raw  
train:confusion>log.txt
```

خروجی با دقت ۹۰٪ بدست آمد (هم برای داده‌ی test و train)

label	0	1	total
0 emam	298	9	307
1 shah	42	45	87

فایل log.txt را که در بالا ایجاد شده را به برنامه‌ی raw.py که نوشتیم دادیم و داده‌هایی که اشتباه تشخیص داده شده بودند را چاپ کردیم و سعی کردیم بهترین feature ها را پیدا کنیم ولی feature بیشتری متأسفانه پیدا نشد.