

## گزارش Naive Bayes :

در این قسمت ابتدا همه‌ی داده‌ها را مرتب و nomralize کرده و سپس تماما tokenize شدند. سپس با استفاده از یک سری دستورات داده را برای train و test جدا کردیم. نتیجه به صورت مقابل شد:

recall:

sohrab: 0.7465753424657534

moshiri: 0.8816568047337278

precision:

sohrab: 0.8449612403100775

moshiri: 0.8010752688172043

برای بهبود نتایج سعی کردم که داده‌ها را stem کنم و دوباره تست کنم اما نتیجه‌ی بهتری صورت نگرفت. (از smoothing در کد استفاده شده)

## گزارش VowPal Wabbit :

در این قسمت از فایل‌های test و train مرحله‌ی قبل استفاده شده تا آن‌ها را به صورت فومت مورد قبول vowpal عوض کنیم. برای کلاس بندی کردن از ۲ عدد 1 و -1 استفاده شده.

سپس فایل train را به دستور زیر اجرا می‌کنیم:

```
vw -d vowpal_train.txt --passes 10 -f predictor.vw --ngram 1 --loss_function quantile -c
```

و بعد فایل test را با دستور زیر اجرا می‌کنیم:

```
vw -d vowpal_test.txt -t -i predictor.vw -p prediction.txt
```

که خروجی مورد نظر در prediction.txt ذخیره می‌شود.

سپس از خروجی استفاده کرده و accuracy, precision, recall را از آن به دست می‌آوریم که نتایج به صورت زیر است:

sohrab results:

sohrab accuracy:0.3900709219858156

sohrab precision:0.7801418439716312

sohrab recall:0.8461538461538461

moshiri results:  
moshiri accuracy:0.42907801418439717  
moshiri precision:0.8581560283687943  
moshiri recall:0.7960526315789473

همان طور که مشاهده می شود precision سهراب بدتر ولی precision مشیری بهتر شده و recall نیز دقیقاً برعکس ( نتایج مربوط به 1-gram می باشد و برای 2-gram بدتر شد )

هومن مهرآفرین ۹۴۵۲۱۲۲۵