

## گزارش فاز دوم پروژه WordCloud :

با استفاده از Naïve Bayes همه ی کلمات در ۳ دیکشنری طبقه بندی میشوند . دیکشنری کلمات پاپ و دیکشنری کلمات رپ و دیکشنری تمامی کلمات و احتمال وجود هر کلمه در هر کلاس محاسبه می گردد. یعنی احتمال وجود یک کلمه در دیکشنری رپ و دیکشنری پاپ وجود دارد. سپس برای بخش test احتمال هر کلمه که به آن بر می خوریم از ۲ کلاس خارج شده و در احتمال کلمات دیگر داده تست ضرب میشود. مجموع لگاریتم های هر کلاس را با یکدیگر مقایسه می کنیم و مقدار هر کدام که بیشتر شد به این معناست که داده ی تست مربوط به آن کلاس است. برای vowpal wabbit کلمات را استخراج می کنیم و کلمات هر کلاس را در یک خط می گذاریم و آن ها را برای ۲ کلاس یکی در میان نوشته و از هم متمایز می کنیم ( هم برای train.txt و هم برای test.txt ) با استفاده از دستورات :

```
+vw -d train.txt -c -f predictor.vw  
vw -d test.txt -t -i predictor.vw -p prediction.txt
```

خروجی در فایل prediction.txt ایجاد میشود.