Pooya jahangiri
98243076

Project 1 AI

ابتدا دیتای train را در یک dataframe میگیریم و کلمات و کاراکترهای بدردنخور را حذف میکنیم (کلمات کمتر از 3 حرف و '!' و ...)(در اینجا کلمات حذف نشده)
سپس که یک لیست از کلمات مورد نیاز هر لیست داریم ، آنها را برمیداریم و با توجه به لیبل آن ها بهشان یک شانس اختصاص می دهیم (probability) که با دیدن آن کلمه، آن لیبل جواب ما هست(با کمک الگوریتم نایبو بایس)
سپس میایم روی دیتای test، در این مرحله دیتا را نگاه میکنیم و هر کلمه در هر query را که میبینیم، شانس آمدن آن کلمه در هر لیبل را میبینیم و هرکدام که شانس بیشتری داشت آن جمله را به آن لیبل نسبت می دهیم، در نهایت هر کدام از کلمات با شانس بیشتری به یک لیبل نسبت داده شد ، آن query هم به آن لیبل نسبت میدهیم.
نمونه خروجی csv :

	id	label
0	0	4
1	1	1
2	2	4
3	3	2
4	4	2
...	...	...
757	757	2
758	758	2
759	759	1
760	760	5
761	761	1