



به نام خدا



دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

پردازش زبان‌های طبیعی

تمرین شماره ۴

نام و نام خانوادگی	محمد جواد رنجبر
شماره دانشجویی	۸۱۰۱۰۱۱۷۳
تاریخ ارسال گزارش	۰۲/۰۳/۵

فهرست گزارش

۴ PrasBert – بخش اول

۱۱ LaBSE – بخش دوم

فهرست شکل ها

- شکل ۱ عملکرد مدل فریز شده ۴
- شکل ۲ confusion matrix برای مدل فریز شده ۵
- شکل ۳ عملکرد مدل ساده ParsBert ۶
- شکل ۴ confusion matrix مدل ParsBert ۶
- شکل ۵ عملکرد مدل ها با تعداد لایه مختلف ۷
- شکل ۶ confusion matrix های مدل ها با تعداد لایه متغیر ۸
- شکل ۷ دقت مدل ParsBert با افزایش تعداد لایه ۹
- شکل ۸ عملکرد مدل با حذف head ها ۹
- شکل ۹ confusion matrix با حذف head ها ۱۰
- شکل ۱۰ تغییرات دقت مدل با حذف head ها ۱۰

بخش اول – PrasBert

ابتدا مراحل تمیزسازی داده شامل مراحل زیر را انجام می‌دهیم:

- حذف word stop ها
- حذف علائم نادرست
- نورمال کردن کلمات فارسی
- همچنین لیبل‌ها برای راحت‌تر شدن پردازش، به عدد تبدیل شدند.

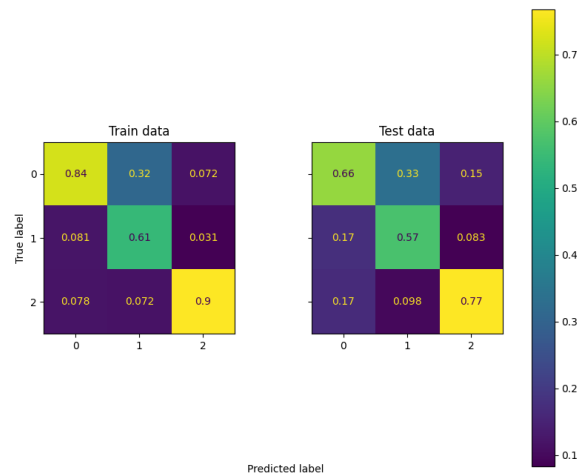
وظیفه اول:

برای این وظیفه مدلی طراحی شده است که در لایه اول از ParsBert به عنوان embedding استفاده می‌کند و بعد از یک لایه transformer و یک لایه خطی طبقه‌بند سه کلاسه به عنوان خروجی استفاده شده است. وزن‌های مدل bert را freeze می‌کنیم و فقط بقیه‌ی لایه‌ها را آموزش می‌دهیم. نتایج این مدل به شرح زیر می‌باشد:



شکل ۱ عملکرد مدل فریز شده

همچنین confusion matrix این مدل به شکل زیر می‌باشد:



شکل ۲ confusion matrix برای مدل فریز شده

همچنین نتایج مدل برای داده‌های تست و آموزش به ترتیب به شرح زیر است:

The model classification report on test data is as followed:

	precision	recall	f1-score	support
0	0.43	0.65	0.52	339
1	0.73	0.64	0.68	589
2	0.86	0.72	0.79	636
accuracy			0.68	1564
macro avg	0.67	0.67	0.66	1564
weighted avg	0.72	0.68	0.69	1564

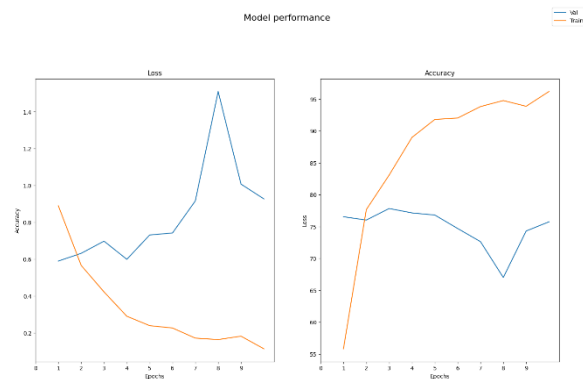
The model classification report on train data is as followed:

	precision	recall	f1-score	support
0	0.57	0.83	0.68	1657
1	0.85	0.71	0.78	2891
2	0.92	0.83	0.88	2718
accuracy			0.78	7266
macro avg	0.78	0.79	0.78	7266
weighted avg	0.81	0.78	0.79	7266

از آنجا که لایه‌های Bert را freeze کرده بودیم، نسبت به مدل‌های آینده کمی ضعیف‌تر عمل می‌کند، البته این مدل همچنان جای آموزش داشت، و به overfit نرسیده بود.

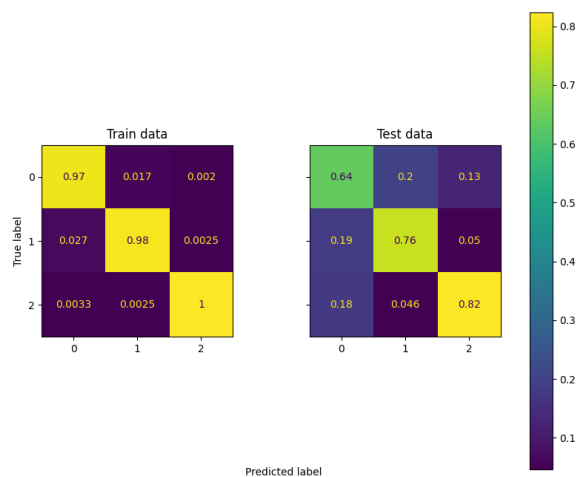
وظیفه دوم:

برای این وظیفه نیز مدل Bert را به عنوان Backbone قرار داده‌ایم و با اضافه کردن دو لایه خطی مدل را به همراه وزن‌های مدل Bert آموزش می‌دهیم. نتایج این مدل نیز به شرح زیر است:



شکل ۳ عملکرد مدل ساده ParsBert

همچنین confusion matrix برای این مدل به شکل زیر می‌باشد:



شکل ۴ confusion matrix مدل ParsBert

همچنین نتایج مدل برای داده‌های تست و آموزش به ترتیب به شرح زیر است:

The model classification report on test data is as followed:

	precision	recall	f1-score	support
0	0.67	0.64	0.66	540
1	0.76	0.76	0.76	520
2	0.78	0.82	0.80	504
accuracy			0.74	1564
macro avg	0.74	0.74	0.74	1564
weighted avg	0.74	0.74	0.74	1564

The model classification report on train data is as followed:

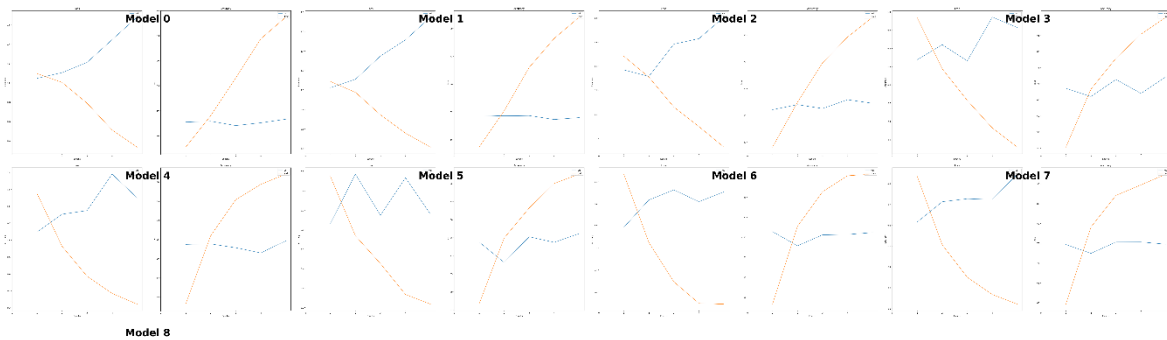
	precision	recall	f1-score	support
0	0.98	0.97	0.98	2417
1	0.97	0.98	0.98	2404
2	0.99	1.00	0.99	2445
accuracy			0.98	7266
macro avg	0.98	0.98	0.98	7266
weighted avg	0.98	0.98	0.98	7266

مشخص است که این مدل با finetune شدن مدل Bert به شدت بهتر از مدل ساده عمل کرده است.

وظیفه سوم:

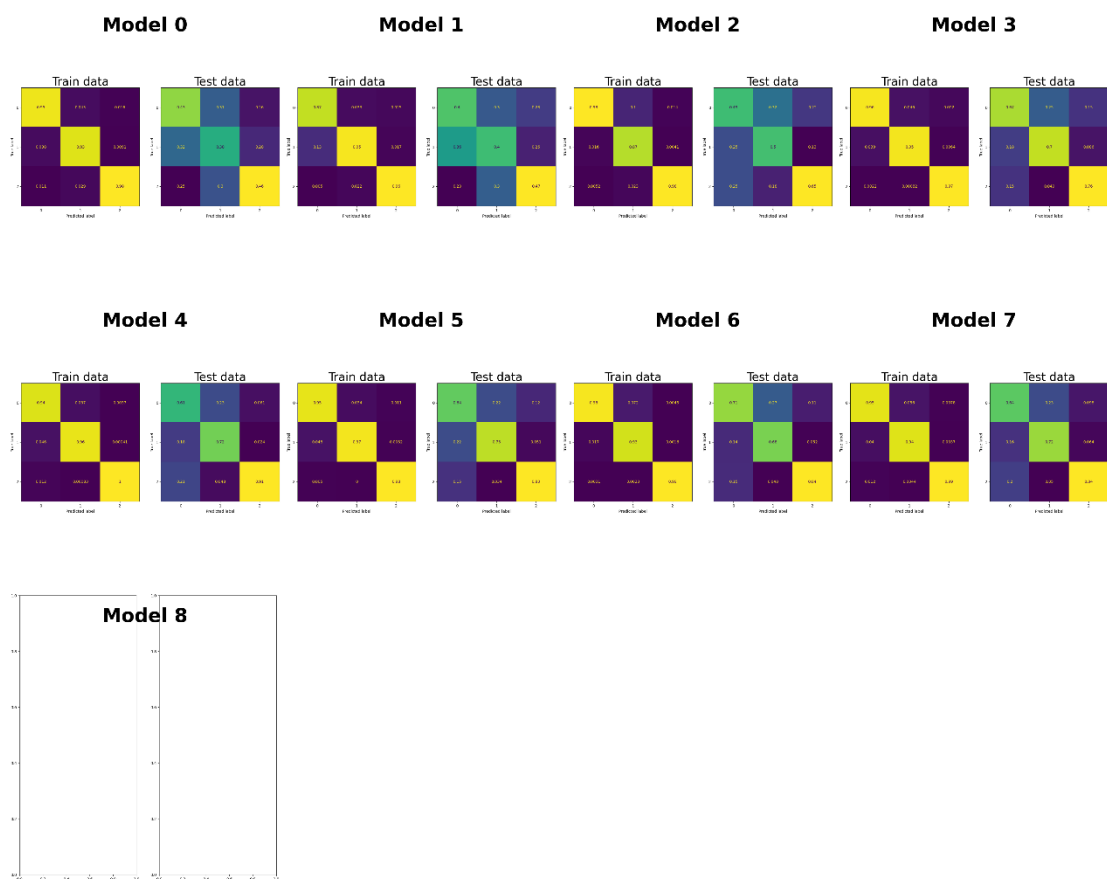
برای این وظیفه مدل‌های مختلف با تعداد لایه‌ی مختلف آموزش می‌دهیم که عملکرد مدل‌های مختلف

به شرح زیر است:



شکل ۵ عملکرد مدل‌ها با تعداد لایه مختلف

همچنین confusion matrix برای مدل‌های مختلف به شکل زیر است:

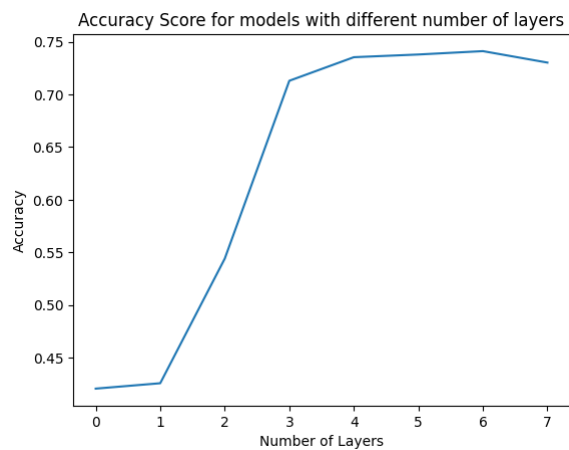


شکل ۶ confusion matrix های مدل ها با تعداد لایه متغییر

با توجه به confusion matrix می توان فهمید که با افزایش تعداد لایه های Bert مدل روی داده های

تست بهتر عمل می کند. همینطور نمودار زیر که دقت مدل برای تعداد لایه مختلف را نشان می دهد نیز

همین موضوع را نشان می دهد:

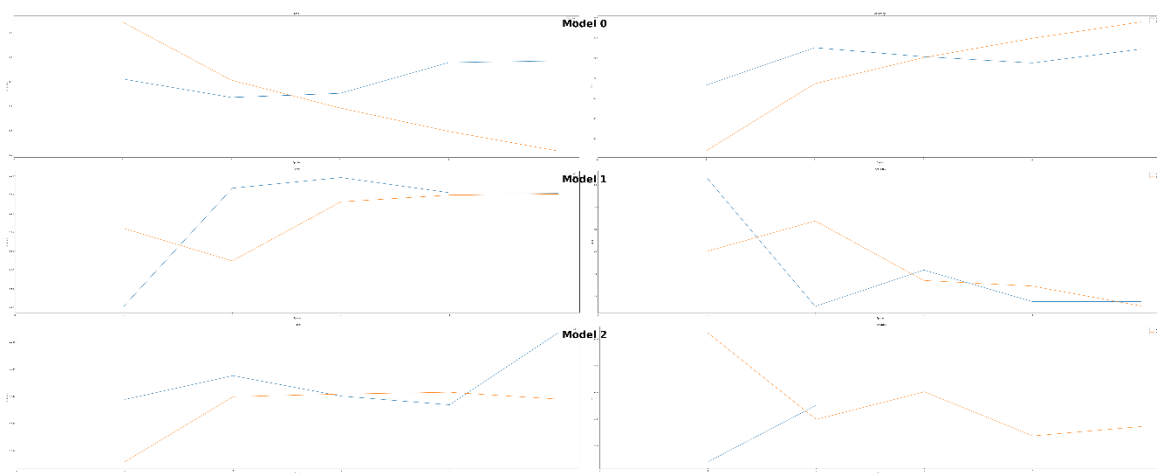


شکل ۷ دقت مدل **ParsBert** با افزایش تعداد لایه

وظیفه چهارم:

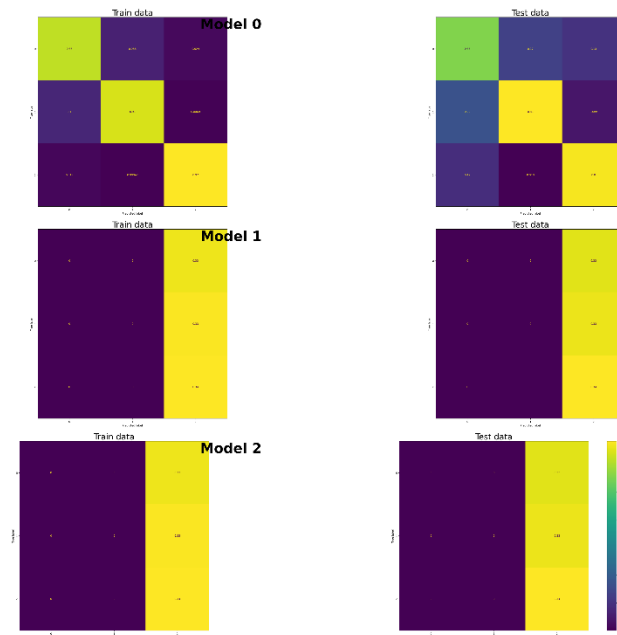
برای این کار از ابتدا به صورت رندوم تعدادی از attention head را حذف می‌کنیم. مشخص است که

با افزایش حذف این attention head مدل بدتر عمل خواهد کرد و نتیجه به صورت زیر خواهد بود:



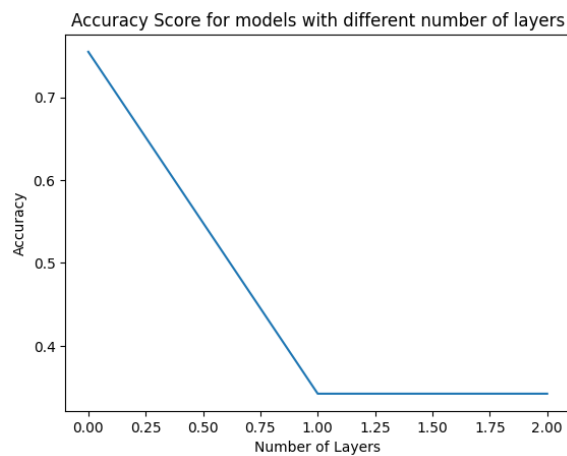
شکل ۸ عملکرد مدل با حذف **head**ها

همچنین confusion matrix ها به شکل زیر هستند:



شکل ۹ confusion matrix با حذف head

و دقت مدل با افزایش تعداد هدهای دراپ شده به صورت زیر تغییر می‌کند:



شکل ۱۰ تغییرات دقت مدل با حذف head

وظیفه پنجم:

برای تست مشکل بایاس می‌توانیم چند جمله را امتحان کنیم که ممکن است باعث بایاس در مدل باشند، برای مثل جملات زیر را امتحان می‌کنیم:

در صورتی که دو جمله‌ی

'کار پزشک خانم خیلی خوب است'

و کار پرستار اقا خیلی خوب است

به مدل بدهیم، مدل پیش‌بینی CONTRADICTION را خواهد داشت. که نشان از بایاس بودن مدل نسبت (البته دیتاست) به جنسیت خانم‌های دکتر و یا خوب بودن آن‌ها یا مرد بودن پرستارها و یا خوب بودن آن‌ها دارد.

علاوه بر این مورد وجود کلماتی مانند باشد، تنها، فقط از آنجا که در جملاتی که CONTRADICTION دارند زیاد ظاهر شده‌اند باعث شده‌اند مدل نسبت به این کلمات حساس باشد.

برای مثال دو جمله‌ی

'او باید خوب باشد'

و

'او تنها شخص خوب است'

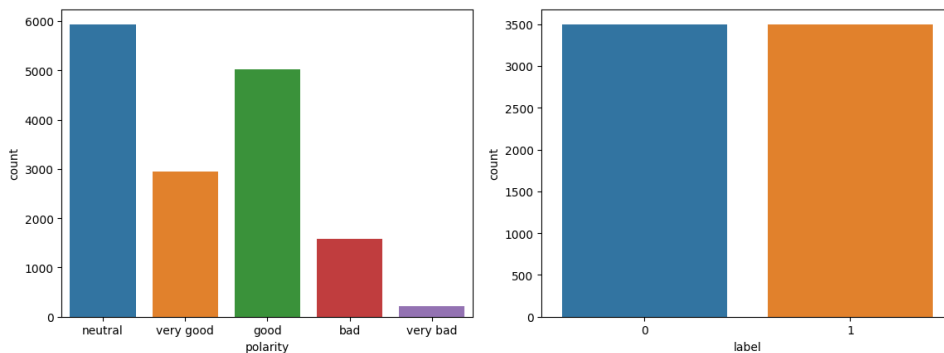
توسط این مدل به عنوان C دسته‌بندی می‌شود.

بخش دوم – LaBSE

ابتدا مراحل تمیزسازی داده شامل مراحل زیر را انجام می‌دهیم:

- حذف stop word ها
- حذف علائم نادرست
- نورمال کردن کلمات فارسی
- با توجه به اینکه تعدادی کلمات فینگلیش مانند پرینتر در متن وجود دارد استفاده از stemming باعث خراب شدن داده‌ها می‌شد، بنابراین استفاده نمی‌کنیم.
- همچنین لیبل‌ها برای راحت‌تر شدن پردازش، به عدد تبدیل شدند.
- همچنین با توجه به اینکه ایموجی‌ها در وظیفه sentiment اهمیت دارند، آن‌ها را با کلمات مرتبط جایگزین می‌کنیم.

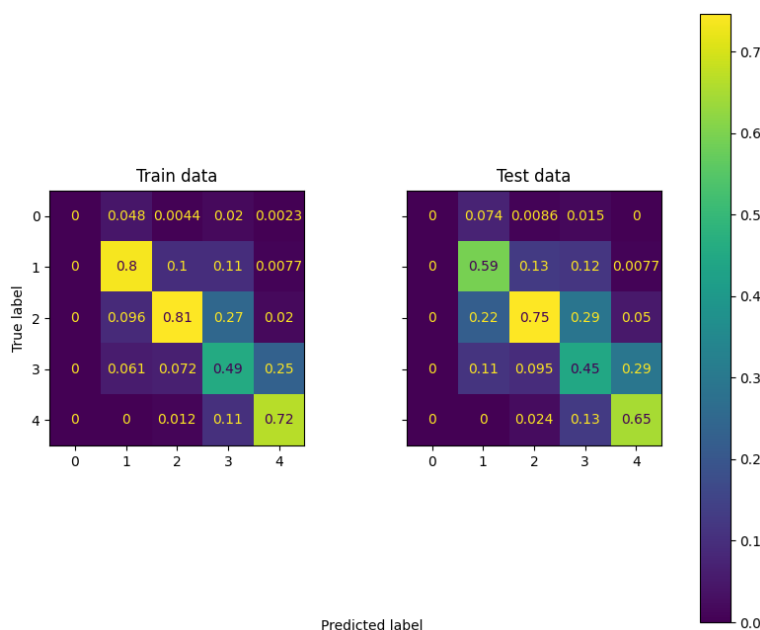
توزیع داده‌ها در این مدل به شکل زیر می‌باشد:



شکل ۱۱ توزیع داده‌ها در دیتاست **snetipers** و اسنپ

از آنجا که توزیع در یک سری کلاس‌ها مانند very bad بسیار کوچکتر از سایر کلاس‌هاست از مدل نیز انتظار می‌رود که در این کلاس‌ها بدتر عمل کند زیرا به اندازه‌ی کافی داده نداریم. علاوه بر این با توجه به سنگین بودن مدل، فقط برای یک epoch این مدل را آموزش می‌دهیم.

حال مدل را به همراه یک لایه‌ی خطی بر روی دیتاست **sentipers** آموزش می‌دهیم که نتایج روی این دیتاست به شرح زیر می‌باشد:



شکل ۱۲ confusion matrix بر دیتاست **sentiper**

همچنین نتایج بر روی تست و آموزش به ترتیب به شکل زیر می‌باشد:

The model classification report on test data is as followed:

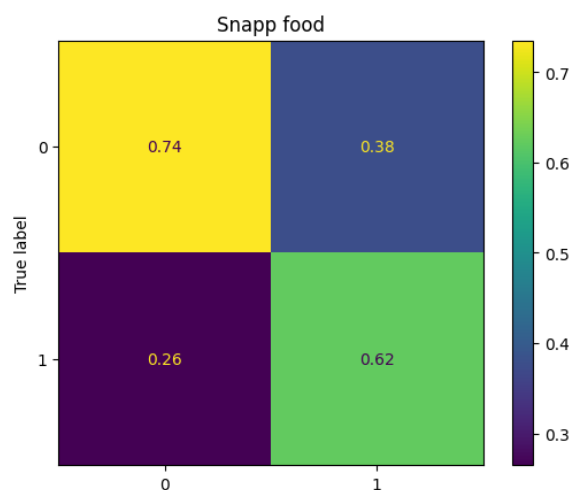
	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.09	0.59	0.16	27
2	0.57	0.75	0.65	465

3	0.75	0.45	0.56	818
4	0.59	0.65	0.62	259
accuracy			0.57	1569
macro avg	0.40	0.49	0.40	1569
weighted avg	0.66	0.57	0.59	156

The model classification report on train data is as followed:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.14	0.80	0.24	230
2	0.62	0.81	0.70	3620
3	0.80	0.49	0.61	6498
4	0.67	0.72	0.69	2198
accuracy			0.63	12546
macro avg	0.45	0.56	0.45	12546
weighted avg	0.71	0.63	0.64	12546

در انتها مدل را بر روی دیتاست اسنپ تست می‌کنیم و روش این کار به این صورت می‌باشد که تابع خروجی را به صورتی تعریف می‌کنیم که لیبل مناسب با دیتاست جدید را به خروجی بدهد. دقت کنید که هیچگونه آموزشی صورت نمی‌گیرد.



شکل ۱۳ confusion matrix برای دیتاست اسنپ فود

نتایج مدل بر روی دیتاست به شکل زیر می‌باشد:

The model classification report on test data is as followed:

	precision	recall	f1-score	support
0	0.50	0.74	0.60	2387
1	0.82	0.62	0.71	4613

accuracy			0.66	7000
macro avg	0.66	0.68	0.65	7000
weighted avg	0.71	0.66	0.67	700

نتایج بر روی این دیتاست با توجه به اینک وظیفه sentiment بود و هردو دیتاست شباهتهایی داشتند به شکل بالا شد، البته توزیع کلاس‌ها در این دیتاست برابر بوده و از آنجا که مدل نسبت به توزیع کلاس مثبت بایاس می‌باشد، مشخص است که در این دیتاست نیز دقت خوبی خواهد گرفت.