

در این تمرین کلمات و جملات مهم متن با استفاده از تجزیه نامنفی ماتریس به دستآمده است.

در ابتدای کد **stop word** ها از جملات حذف شده اند تا دقت مدل بالا تر برود.

در بخش اول تمرین خواسته شده که این کار با تجزیه **svd** انجام شود و همانطور که در جزوه گفته شد این کار معادل این است که تقریب رتبه 1 ماتریس را به دست آوریم و این کار در انتهای نوت بوک انجام شده. در بخش دوم تمرین کلمات کلیدی با استفاده از تقریب رتبه **k** ام انجام شده که در ابتدای نوت بوک آمده است. البته در جروه از ماتریس های هاوس هولدر استفاده شده تا جملات مشابه کمتر انتخاب شوند اما متاسفانه من کدی برای محاسبه ماتریس هاوس هولدر پیدا نکردم و همانطور که پیداست جملات و کلمات مشابه انتخاب شده اند.

تحلیل جواب:

همانطور که انتظار میرفت تقریب رتبه **k** ام دقت بالاتری را به ما میدهد و جملات مهم را بهتر پیدا میکند. همچنین دیده میشود که با استفاده نکردن از ماتریس های هاوس هولدر، کلمات مهم مشابه مثل **computer** و **computing** انتخاب شده اند و برای اینکه کلمات مشابه یکیشان انتخاب شود باید حتما از این ماتریس ها استفاده کنیم.