# تغییر در پیش پردازش داده ها و تمیز کردن مناسب تر دیتاست

برای کلاستر بهتر لغات(گروهبندی) آن ها بهتر است که لغات و علامت های اضافه احتمالی را از آن ها حذف کنیم حروفی مثل اعداد که ربط چندانی به خود اینکه product چی هست ندارند یا علائم نگارشی که باز نبودن آن ها در text ها الزامی است یا باز هم به عنوان مثال کلمات کوچک و بزرگ را همرو یک دست می کنیم که نتیجه بهتر بگیریم.

# تغییر در Vectorization :

تقییر در این قسمت و متد های مختلف می تواند به عملکرد کمک کند در برداری کردن متد های مختلفی هست ولی چون بهترین متد را اعمال کردم ، تغییری در آن ایجاد نمی کنم (در تسک 2 نتیجه انواع بردارسازی را بررسی کردم و می تونید اونجا مشاهده کنید )