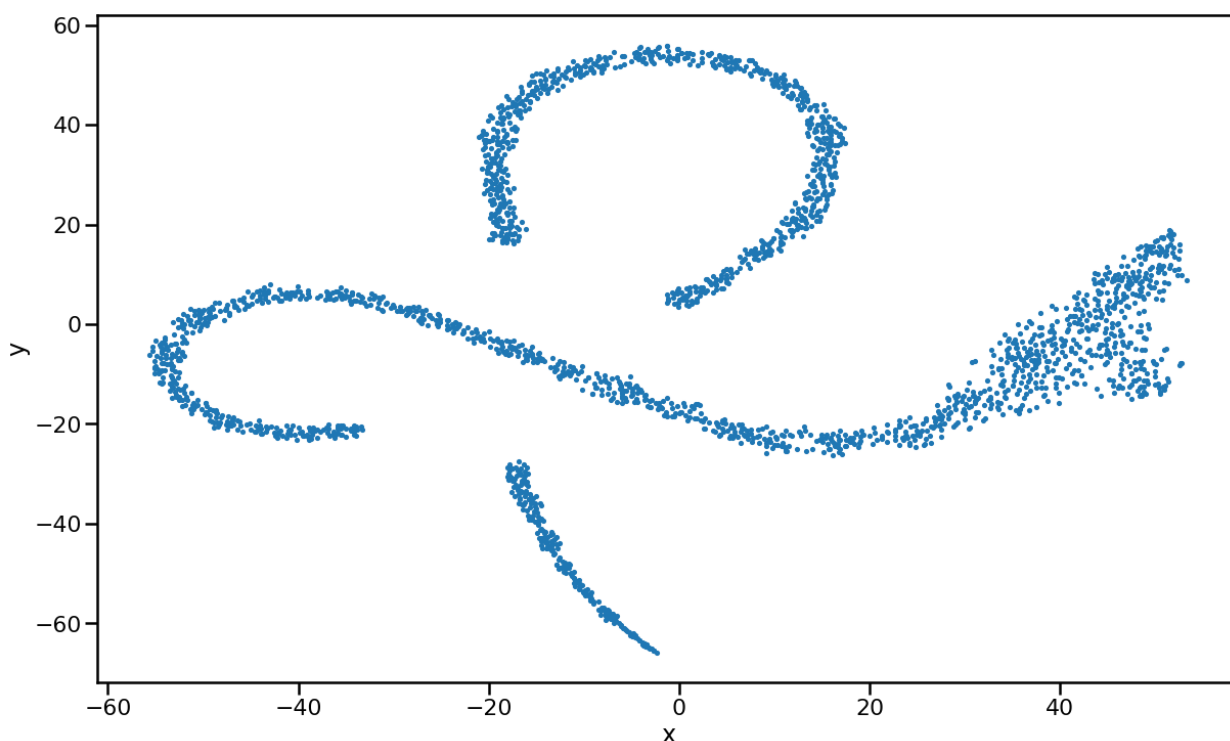


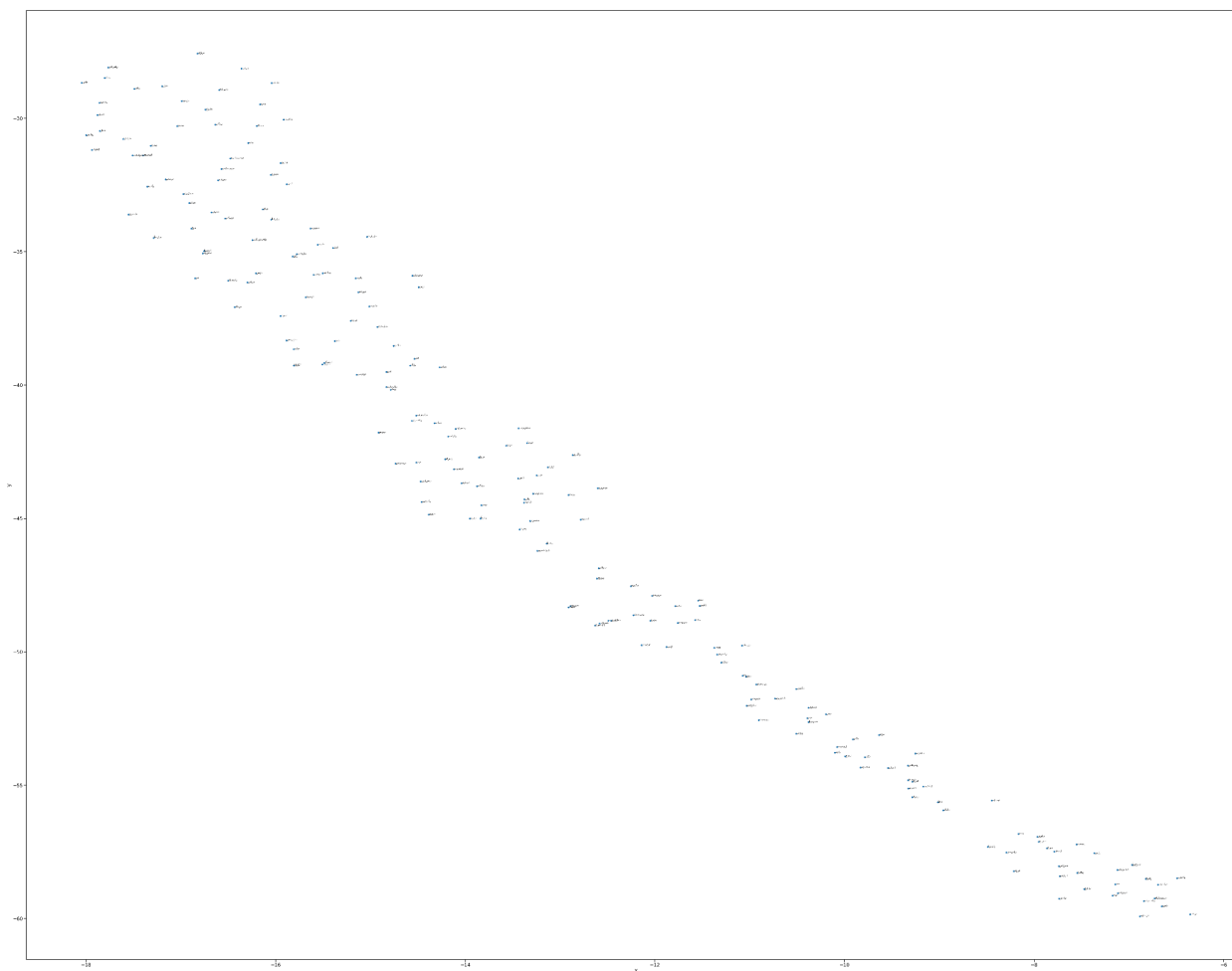
# گزارش خروجی تحلیل متن کتاب دوم مثنوی معنوی با استفاده از متد Word2Vec

تمرین اول درس NLP / استاد قاسمی  
ایمان سلطانی جهانگیرنیا ۲۲۰۷۹۷۰۶۰

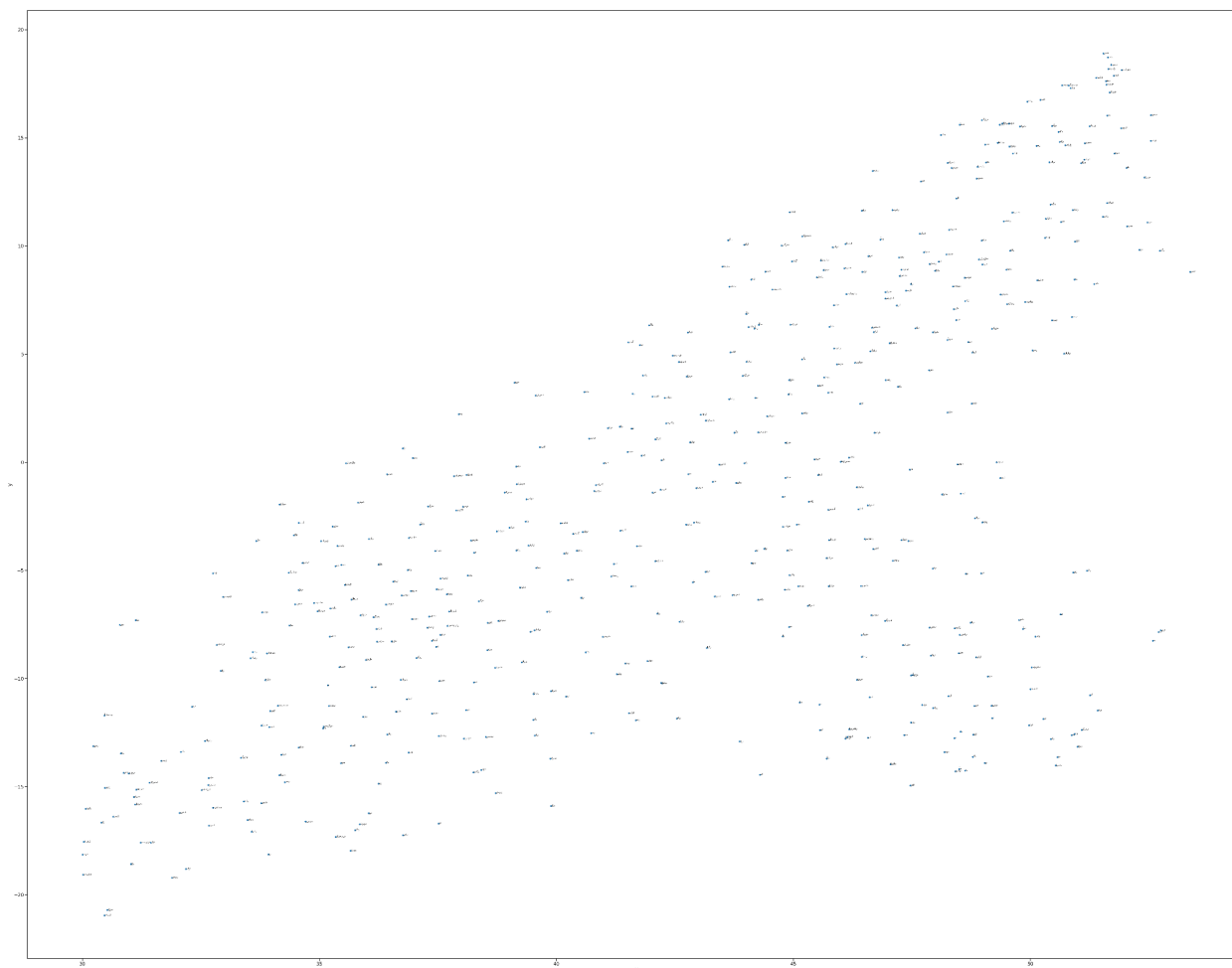
- پس از **Normalize** کردن متن **49,669** کلمه و **3,924** جمله از متن استخراج شد.
- پس از حذف حروف اضافه کلمات را به مدل **Word2Vec** تبدیل شدند
- پس از ساخت مدل **2,402** واژه ی منحصر به فرد استخراج و آماده عملیات **Train** شدند.
- نمودار کلی کلمات به شرح عکس زیر است:



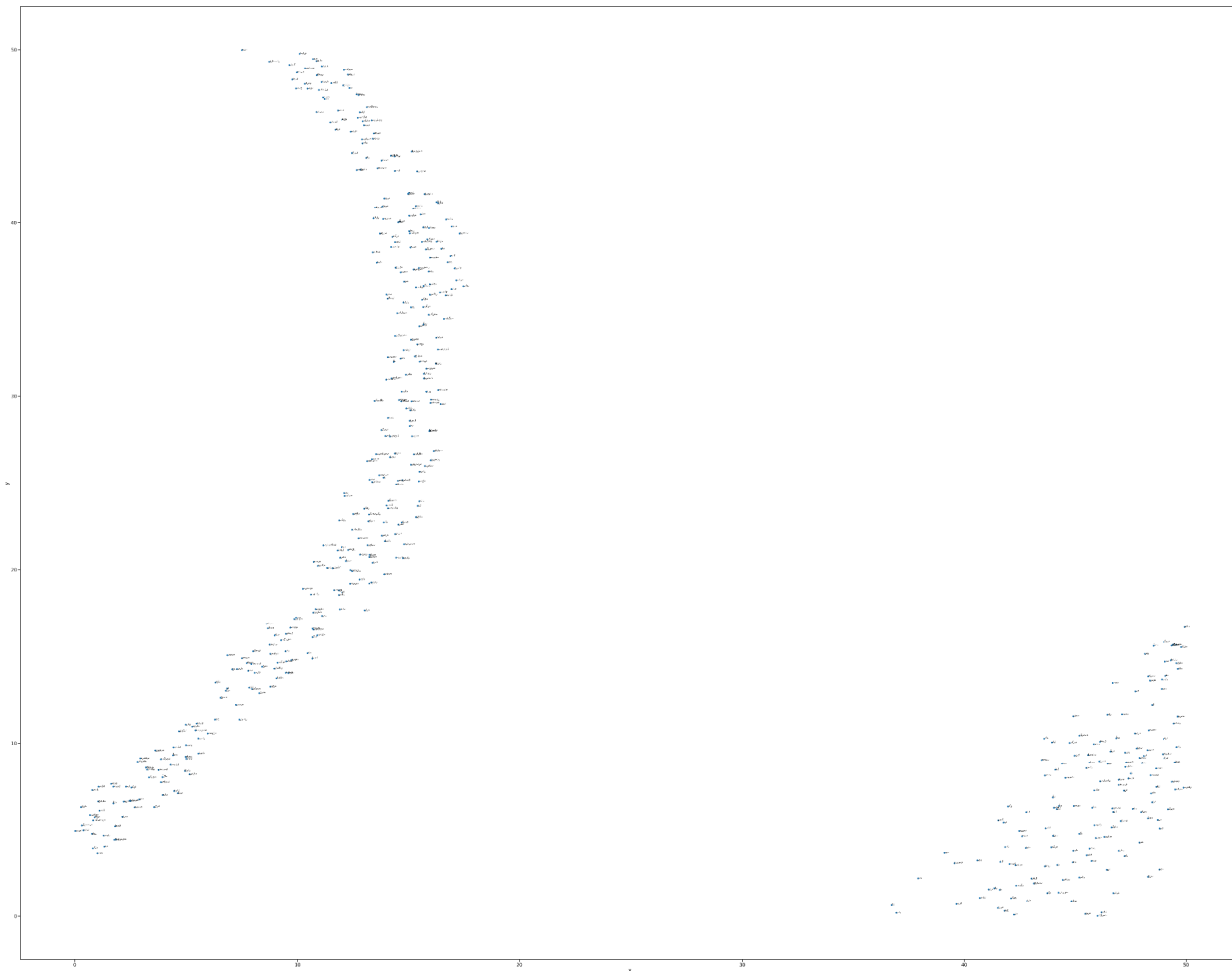
● نمودار جزئی در بازه  $(x(-20, 0), y(-60, -20))$



● نمودار جزئی در بازه  $(x(30, 60), y(-30, 30))$



● نمودار جزئی در بازه  $(x(0, 50), y(0, 50))$



● کلمات مشابه و درصد تشابه آنها با واژه "شمس":

( 'بیا', 0.9998500943183899),  
( 'بگذار', 0.9998422861099243),  
( 'شریف', 0.9998350143432617),  
( 'سیه', 0.999832272529602),  
( 'بدن', 0.9998313188552856),  
( 'مایهی', 0.9998312592506409),  
( 'بوی', 0.9998311400413513),  
( 'دانه', 0.999829888343811),  
( 'نیش', 0.9998290538787842),  
( 'شاهان', 0.9998288154602051)

● کلمات مشابه و درصد تشابه آنها با واژه "دوست":

( 'تنگ', 0.9998462796211243),  
( 'عدو', 0.9998395442962646),  
( 'سرش', 0.999835193157196),  
( 'گنج', 0.9998345375061035),  
( 'خشم', 0.999833345413208),  
( 'کل', 0.9998326897621155),  
( 'هستی', 0.9998317360877991),  
( 'مغز', 0.9998311996459961),  
( 'فعل', 0.9998296499252319),  
( 'ایمن', 0.9998294115066528)

● کلمات مشابه و درصد تشابه آنها با واژه "عشق"

('باطل', 0.9999081492424011),

('روشن', 0.9999029040336609),

('راز', 0.9999027252197266),

('رفته', 0.9999024271965027),

('فزاید', 0.9999021887779236),

('ظاهر', 0.999901294708252),

('برق', 0.9998989701271057),

('کشد', 0.9998989105224609),

('بوی', 0.9998980164527893),

('فراق', 0.999897837638855)

● واژه نا مرتبط در جمله "نور باقی پهلوی دنیای دون شیر صافی پهلوی

جوهای خون" واژه ی "نور" است.

● ارتباط "شمس" با "قرآن" مانند "شریف" با "علم" است.