NLP for House Price Prediction

EPPS7V81

XIAOYAN ZHANG

House Data

	C	D	E	F	G	H	I	J	K	L	M	N
1	ZIP	PRICE	BEDS	BATHS	SQUAREFEET	LOTSIZE	YEARBUIL'	Price_per	MLS	LATITUDE	LONGITUC	HOUSE_DES
2	75007	375000	3	2	1821	8059	1974	206	20193171	33.00057	-96.881	Beautifully updated and well-maintained home on a qu
3	75007	389900	4	2	1706	7841	1979	229	20006334	33.00877	-96.8727	Lovely Carrollton gem with sparkling POOL! Originally I
4	75007	608000	3	3	2721	9191	1986	223	20131908	32.9898	-96.9153	Beautifully updated home with a spectacular park view
5	75007	410000	3	2	1822	7492	1984	225	20141065	32.99164	-96.8944	Multiple offers received. Please submit highest and bes
6	75007	598000	5	3.5	3141	9845	1989	190	20097778	33.00383	-96.9186	Beautiful 5 bed charmer with a POOL in Carrollton's Ti
7	75007	420000	3	2.5	2028	10454	1985	207	20216920	33.02227	-96.8694	Welcome home to this beautiful One Story on a corner
8	75007	395000	3	2.5	1996	4704	2004	198	20000858	33.02249	-96.8874	So many updates in this spectacular Quail Creek townI
9	75007	375000	3	2	1883	9191	1984	199	20050130	33.01856	-96.8782	Amazing opportunity to own a home in desirable Carro
10	75007	377000	3	2	1247	8189	1984	302	20115453	33.01658	-96.874	Welcome to this gorgeous neighborhood! Terrific 3 be
11	75007	575000	4	2.5	2986	5227	2004	193	20036320	33.02151	-96.8898	*MULTIPLE OFFERS RECEIVED PLEASE HAVE ALL
12	75007	499900	4	2.5	2386	8973	1978	210	20001789	32.99148	-96.9167	Gorgeous custom home in Nottingham estates! Newly I
13	75007	310000	3	2	1580	8625	1973	196	20002846	32.99328	-96.8861	MULTIPLE OFFERS RECEIVED: Best and final by Su
14	75007	305000	3	2	1862	8494	1977	164	14749254	33.01501	-96.8905	*PRIME LOCATION* Located Very close to Shops, Re
15	75007	450000	4	2.5	2338	5881	1995	192	20026834	33.0121	-96.9287	Offer Deadline Monday 4.18.22 at noon. Here is a cha
16	75007	540000	4	3	2231	9087	1978	242	20123474	32.9945	-96.9108	Meticulously maintained home in the sought-after Notti
17	75007	439000	4	2.5	2442	7536	1991	180	20021308	32.9987	-96.919	Beautiful one story home in Timbercreek Estates. Waii

House Description

About This Home

Perfectly updated home on greenbelt lot within Panther Creek Estates! Exquisitely and tastefully updated throughout. Entry leads to stacked living & dining room, and study with French doors. Kitchen boasts a large island, gas range and granite countertops and opens to large family room with a beautiful gas fireplace. Masterbed has a view of the greenbelt, dual vanities, separate shower, garden tub and custom walk-in closet. Beautiful backyard with covered patio and flagstone patio. Conveniently located to restaurants, shopping, schools and major highways. Move in ready!

Word2Vec

- Vector representation of the words

- Neural network is used

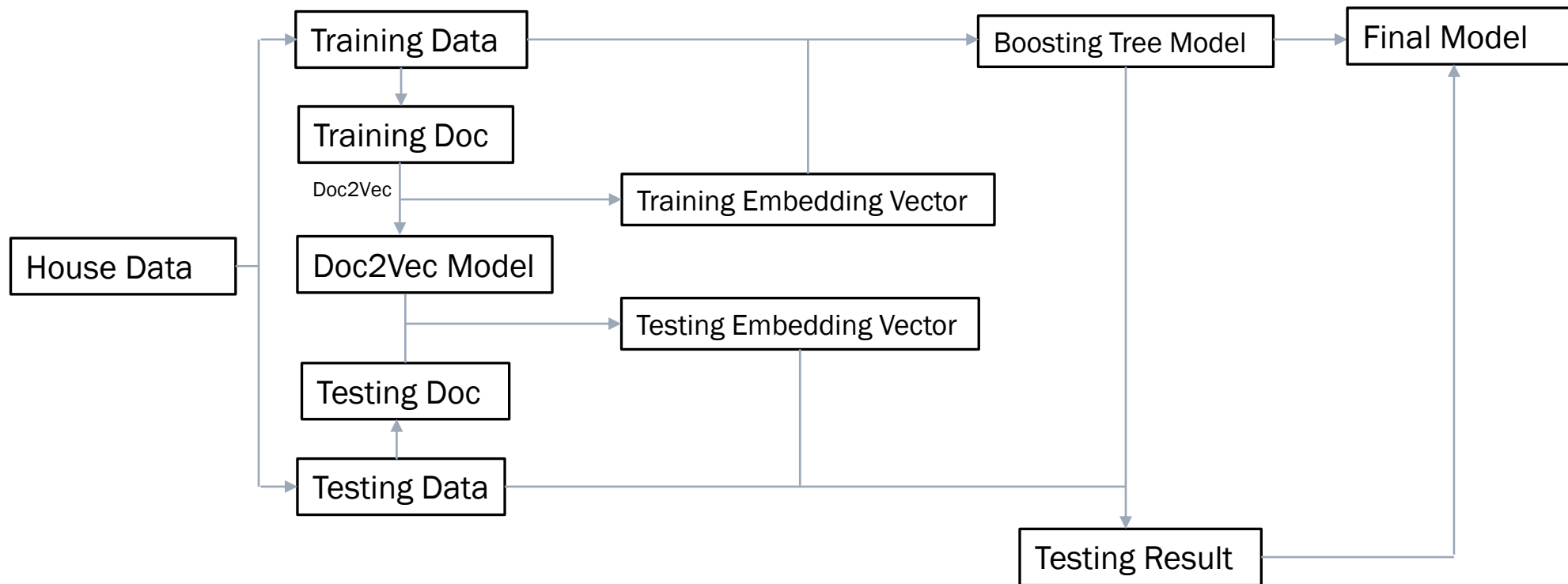- Semantic and syntactic similarity is kept

Doc2Vec

- Extension of Word2Vec

- Summarization of words in Doc

Data Cleaning

➢ Delete non-English words

➢ Remove stop words

➢ Impute missing value and outliers

➢ Derive new variables

Workflow

Model Results

	id	vec_1	vec_2	vec_3	vec_4	vec_5	vec_6	vec_7	vec_8	vec_9	...	vec_15
0	20097778	0.020622	-0.013606	-0.011450	-0.011495	-0.004049	0.017316	0.016412	-0.000023	-0.019717	...	-0.002803
1	20216920	0.013002	0.019194	-0.019423	0.008169	0.008670	0.001017	0.012205	-0.017781	-0.011969	...	-0.012313
2	20000858	-0.006250	0.012602	0.019717	0.015397	0.001287	-0.005755	-0.010618	-0.004431	-0.015596	...	-0.019245
3	20001789	-0.009324	-0.014369	-0.012443	-0.003566	-0.004840	0.006837	0.006825	-0.005770	0.006662	...	-0.003836
4	20026834	-0.005841	0.010300	0.008688	0.007839	0.002021	-0.013495	0.012813	-0.019153	0.017067	...	0.012511

Note: The dimension is 24

Model Results

TSNE visualization of nlp embeddings(validation)

Model Results

TSNE visualization of nlp embeddings (sampled)
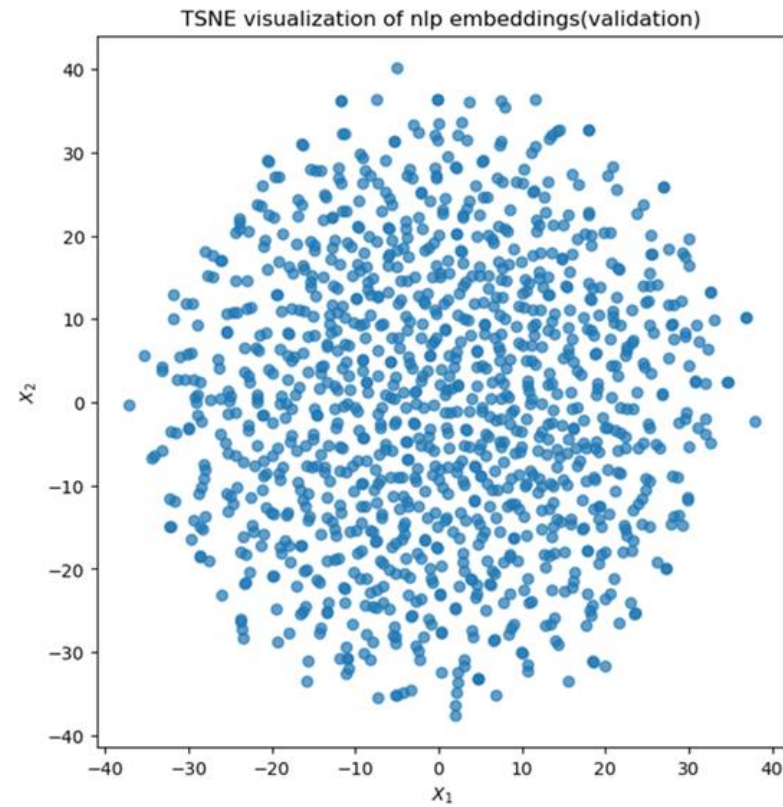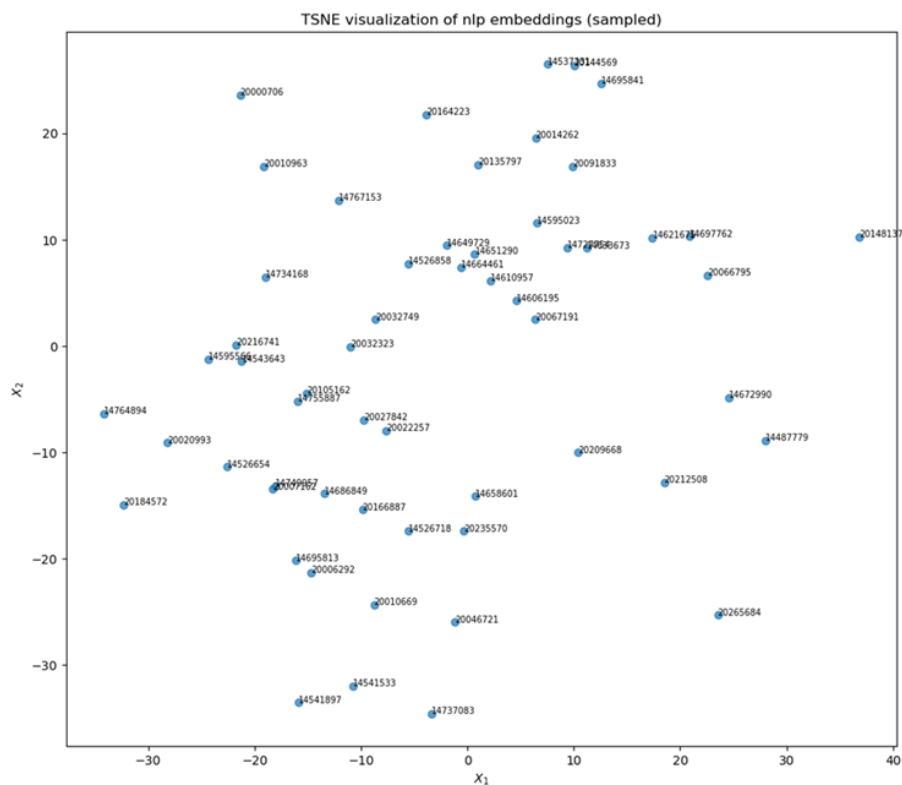
```
id1 = "20216741"
id2 = "14543643"
vec1=valid_vec_dict[id1]
vec2=valid_vec_dict[id2]
similarity = 1 - cosine(vec1, vec2)
similarity
```

0.6146985292434692

```
id1 = "20216741"
id2 = "20265684"
vec1=valid_vec_dict[id1]
vec2=valid_vec_dict[id2]
similarity = 1 - cosine(vec1, vec2)
similarity
```

-0.050728023052215576

Model Results

Multiple Offers received, Highest and Best offer by 12.27 12pm. This Texas beauty is a rare find with its extra-large lot and a fabulous pool located in West Frisco. The floor plan has it all - Primary bedroom plus one office down and 4 Bedrooms up. Each Bedrooms upstairs occupies its own corner of the home for additional privacy. Stunning renovations with trendy updates throughout. High ceilings and tall windows overlook an amazing back yard that is perfect for entertaining friends and family. Two living areas down, 2 LRs up, one room over garage is perfect for a media room or Gym. Recent improvements include 2 recent HVAC systems. roof, roof-gutters, Sidings, All Bathrooms upgraded, stone fireplace, hardwood flooring, fresh paint and many more updates. Please check out the upgrades list! Low HOA fee, few mins to highly rated Frisco elementary. Easy access to DNT, shopping, schools, waterpark. Oversized 3 car garage, backyard oasis and rolling gate which includes long driveway.

This is the ONE you've been waiting for. Fresh paint and kitchen upgrades make it move it ready for you to call home. The large barn door on the game room makes it easy to close off for movie or game nights. The dining room is being used as an office is a great flex space. All bedrooms are upstairs so you can keep a close eye and with large rooms there is plenty of space for adventure. The backyard is where this house really continues the fun. Plenty of room for team parties and gatherings. Please email all questions & offers to LA Elizabeth @ elizabethtxrealtor@gmail.com

Model Results

variable importance
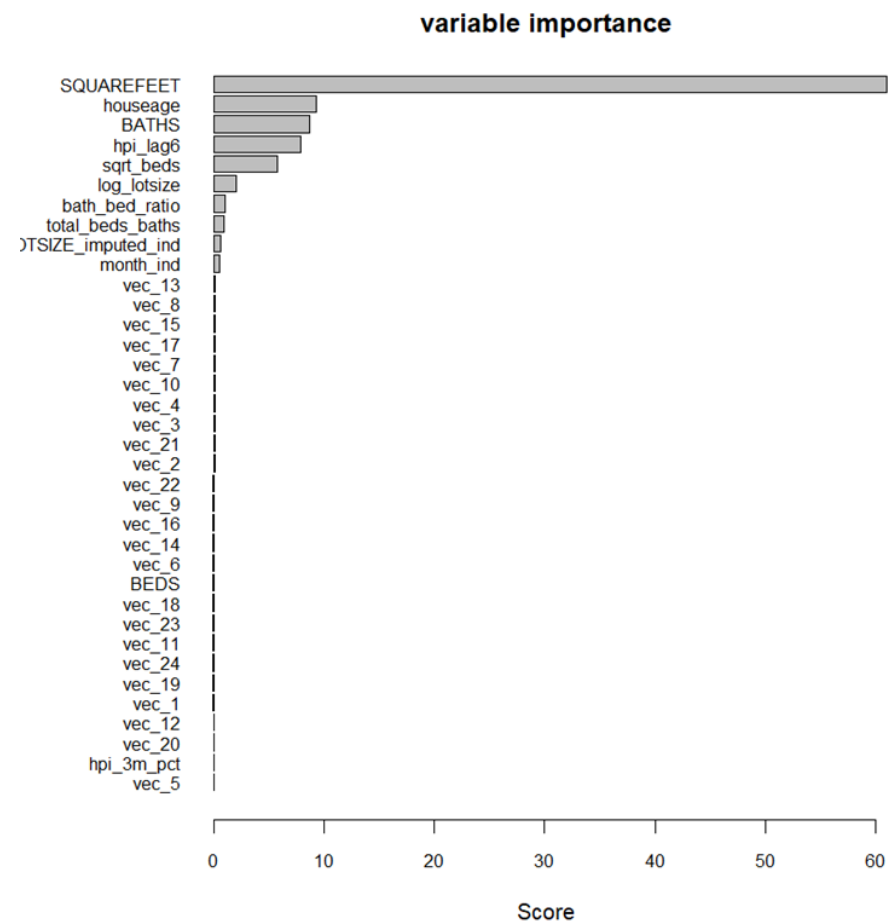
Model Results
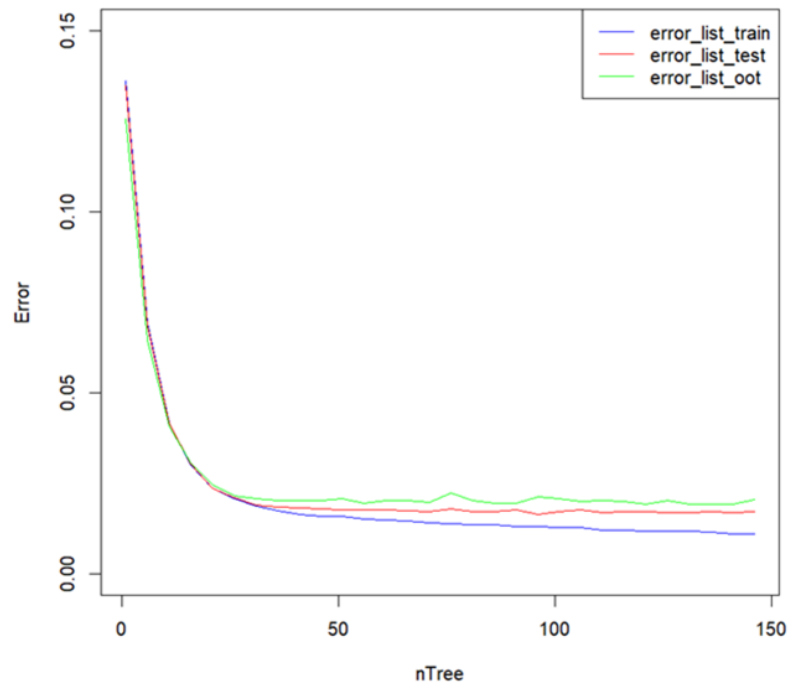
Best model
ntree= 96, learning rate=0.11, max_depth=5
MSE= 0.01726858

Thank You