

Խնդրի դրվածքը՝ հիմնվելով ունեցած տվյալների վրա, կառուցել մոդել, որը առավելագույն ճշտությամբ կկանխատեսի ուսանողը կընդունվի համալսարան թե ոչ

Աշխատանքը կատարվել է Python-ի միջոցով, որպես մոդել օգտագործվել է Deep Neural Networks-ը: Չեն օգտագործվել Python-ի գրականություններ ինչպիսիք են՝ tensorflow, keras և caffe, այլ մոդելը կառուցվել է հիմնվելով դրա մաթեմատիկական սահմանումների վրա:

Մոդելը կառուցվել է քայլերի հետևյալ հաջորդականությամբ՝

1. Initializing the weights and applying forward propagation
2. Loading some of the constants in cache and applying back propagation
3. Updating the parameters
4. Running the code in the main function
5. Designing a prediction function

Ստորև ներկայացված են տվյալները առաջին 5 դիտարկման համար՝

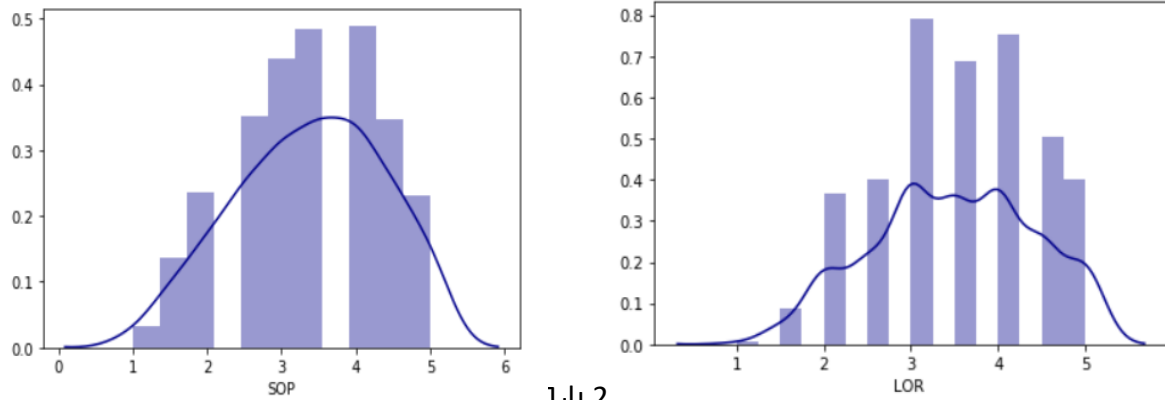
Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit	
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65

Նկ.1

Ունենք 500 ուսանողների տվյալներ (UCLA Graduate Dataset) և փոփոխականներ, որոնք համարվում են կարևոր ուսանողների մագիստրոսական ծրագրերին դիմելու համար:

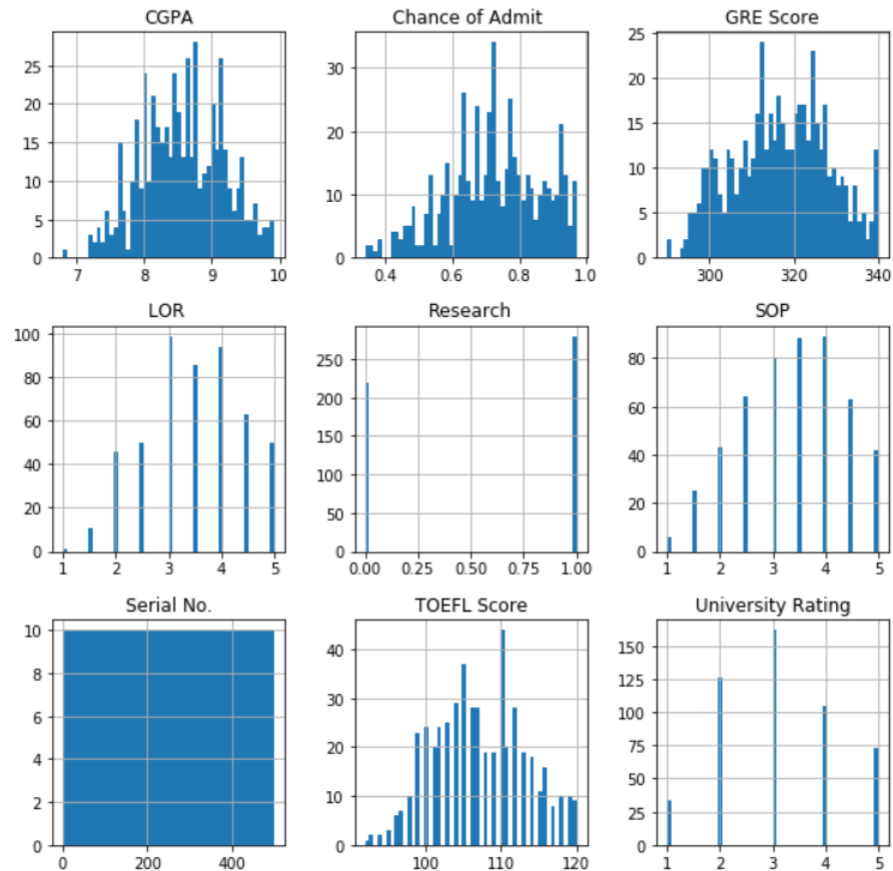
1. GRE Scores (out of 340)
2. TOEFL Scores (out of 120)
3. University Rating (out of 5)
4. Statement of Purpose and Letter of Recommendation Strength (out of 5)
5. Undergraduate GPA (out of 10)
6. Research Experience (either 0 or 1)
7. Chance of Admit (ranging from 0 to 1)

Կատարվել է տվյալների վերլուծություն ըստ որի ուսանողների հիմնական մասը սովորել է 3 վարկանիշ ունեցող համալսարաններում և միայն քչերն են սովորել բարձր վարկանիշ ունեցող համալսարաններում՝ 73 հոգի: Ինչպես նաև պարզվեց որ ուսանողները մեծամասամբ գրել են գերազանց Statement of Purpose (SOP) and Letter of Recommendation (LOR):



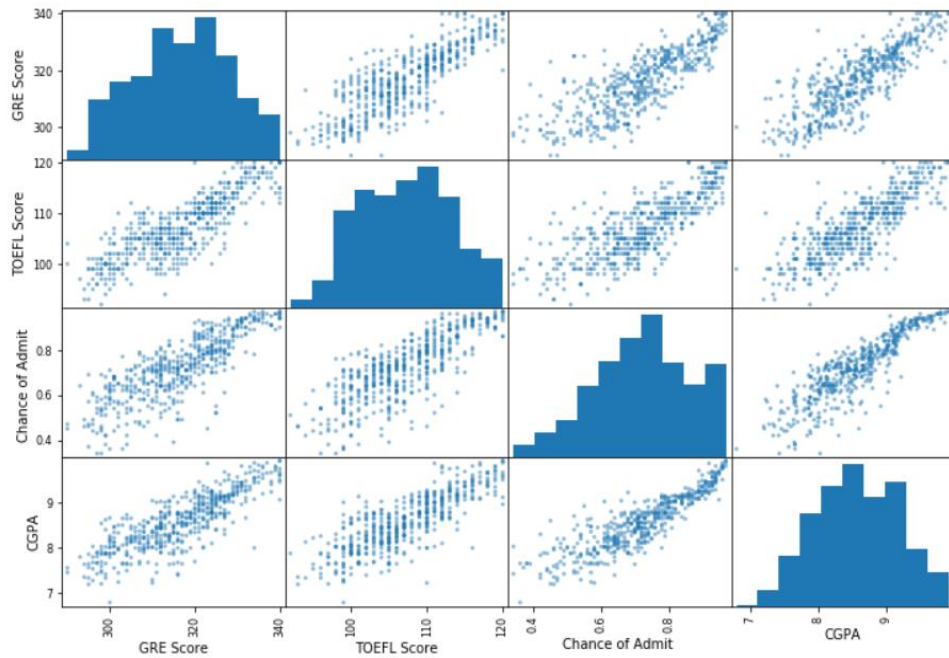
Նկ.2

Փոփոխականների միջև հարաբերությունը հասկանալու համար կառուցվել են ռեգրեսիոն տվյալները ներկայացնող գրաֆիկներ.



Նկ.3

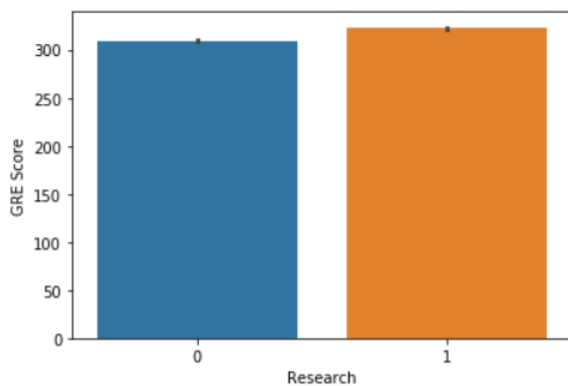
Նկ.3-ում բոլոր փոփոխականները ներկայացված են հիստոգրամների միջոցով, որտեղ երևում է նրանց բաշխվածությունը, ինչը հետագայում կարևոր է ճիշտ վերլուծություններ իրականացնելու համար:



Նկ.4

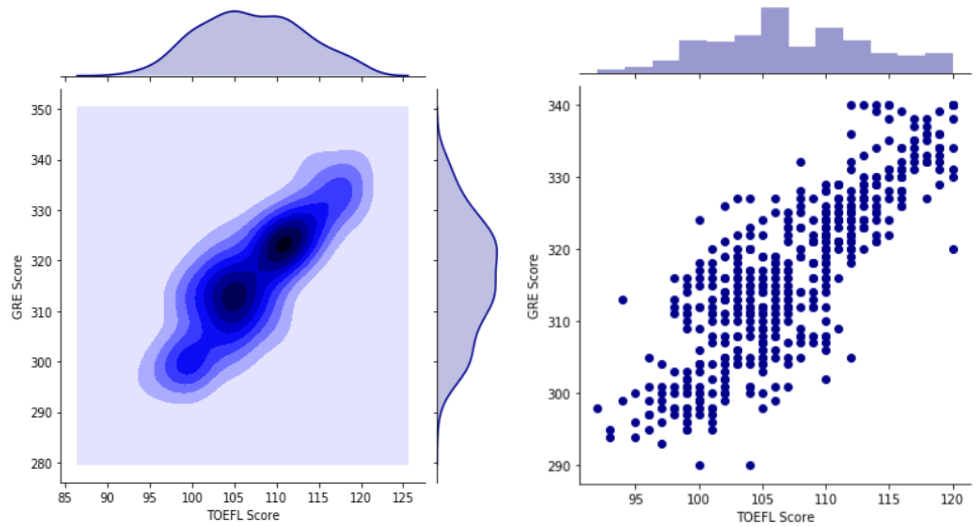
Նկ.4-ում փոփոխականների փոխհարաբերությունն է:

Research և GRE Scores փոփոխականների համար կառուցվել է barplot, որտեղից կարելի է եզրակացնել, որ այն ուսանողները ովքեր հետազոտություններ (Research) են արել, նրանց GRE Score-ը 300-ից բարձր է, ի տարբերություն այն ուսանողների ովքեր չեն արել:



Նկ. 5

Նաև տարբեր վիզուալիզացիաներից հասկացանք, որ եթե GRE Score-ը աճում է, ապա աճում է նաև TOEFL Score-ը:

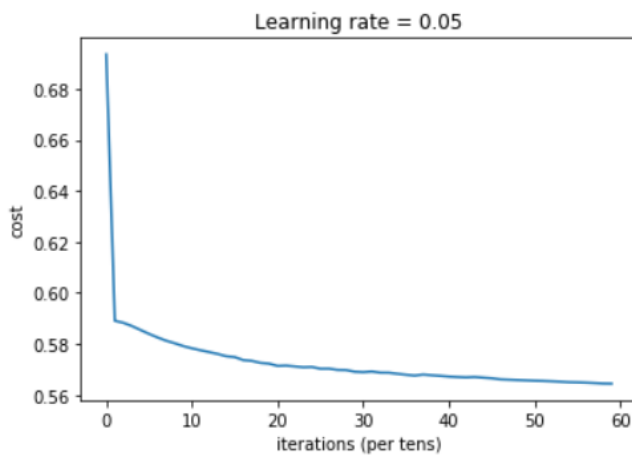


Նկ.6

(Մնացած փոփոխականների համար ևս արված են վիզուալիզացիաներ, որոնք կարող եք գտնել Codes բաժնում)

Հաջորդ քայլում տվյալները բաժանված են test (33%) և train մասերի: Որպես կանխատեսվող փոփոխական վերցված է Chance of Admit փոփոխականը:

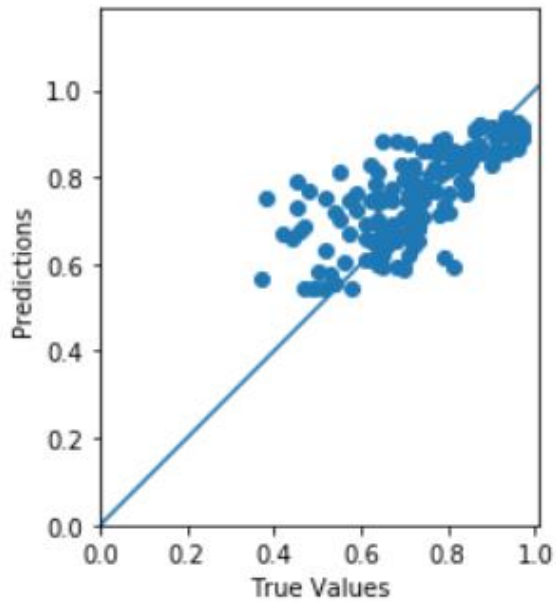
Running the code in the main function -ի կառուցումից հետո ունենում ենք հետևյալ արդյունքը.



Նկ.7

Տեսնում ենք, որ training set error-ը նվազում է խորերացիաների աճին զուգընթաց (տեսնում ենք նաև որ overfit չեն եղել տվյալները): Այս պարագայում մոդելը շատ լավ է աշխատել training set-ի համար և համեմատաբար վատ testing set-ի համար:

Ի վերջո, մոդելի կառուցումից և կանխատեսումը իրականացնելուց հետո ունենք բավականին բարձր արդյունք մոտավորապես 93% ճշտությամբ մեր մոդելը, կարողանում է կանխատեսել ուսանողը կրնդունվի համալսարան թե ոչ՝ ունենալով վերոնշյալ տվյալները:



Նկ.8