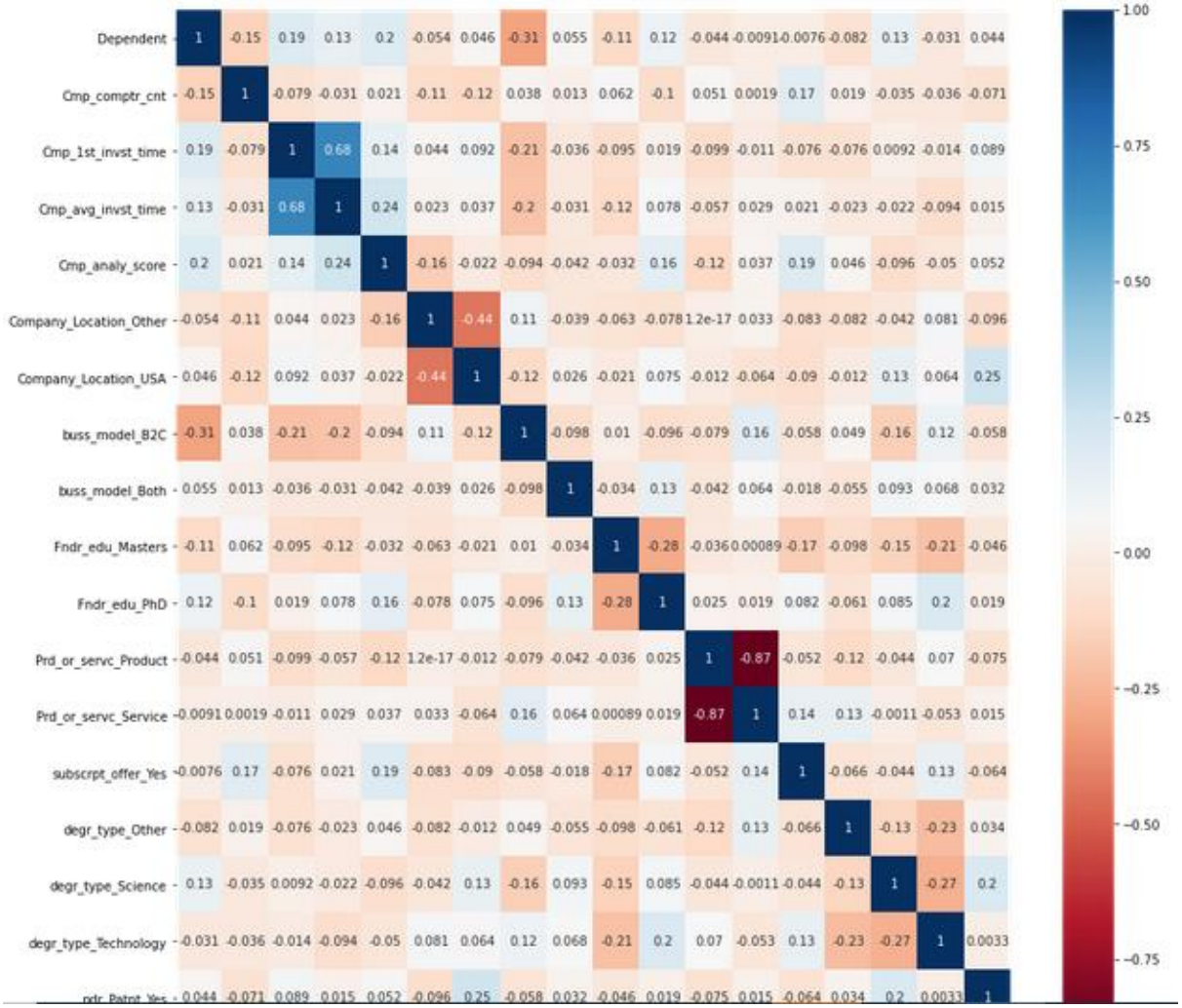
MODELING REPORT

1. **Aim of the project**: use liner model to predict success or failure of a startup
2. **Data set type**: structured data
3. **Data set balance /unbalanced**: data set is balanced. 0=118,1=116
4. **Steps taken to implement the project:**
    1. **Renaming the columns:** the column were renamed for better redabilty , because the name were to big and could not be seen properly on jupyter notebook
    Eg . column 'Founders_previous_company_employee_count' was renamed as fndr_prev_cmp_emp_cnt
    2. **Concatenating train and test:** the dataframe of train and test were concatenated so that similar operations can be performed on the train and test data set. After the operations on the data set is over the data frame will be split to regain train and test dataframes.
    3. **Data standardization:** data standardization results in lower AUC , therefore skipping it.
    4. **Correlation heat map:** the correlation heat map was used to find out the correlation between the dependent and independent variable.
5. 

6. **Feature selection:** for feature selection, univariate and feature importance methods were used, score function was chi2. The top 47 features were selected. As it gave highest AUC.

	Specs	Score
54	ndr_Patnt_Yes	176.000000
35	Cmp_1st_invst_time	49.070470
30	Fndr_Prd_Manag_score	47.670314
9	Comp_rpt_inves	30.627884
24	Fndr_Operat_score	29.114717
6	Comp_advi_cnt	25.060321
32	Fnder_Domain_score	24.921666
41	Cmp_dffclty_obtng_wrkfrc	19.671398
7	Comp_sen_team	19.326834
3	Com_inves_count_seed	18.263916
33	Cmp_incubtn_invst	16.992604
31	Fnder_Sales_score	16.959938
21	Fndr_pblcatn	16.637959
23	Fndr_Entrprnshp_score	16.389931
28	Fndr_ds_score	16.275006
20	Fndr_prof_smlarty	15.920161
27	Fndr_Ldrshp_score	15.669302
4	Com_inves_count_Angel_VC	14.121054
26	Fndr_Marktng_score	13.662790
16	Fndr_Indus_exp	11.850190
52	degr_type_Science	8.374914
34	Cmp_comptr_cnt	7.449213
38	Cmp_crwdfndng	4.284091
39	Cmp_big_data	4.284091
29	Fndr_Buss_Strtgy_score	4.229974
43	Company_Location_USA	3.387717
0	Company_fund	2.822492
17	Fndr_uni_qual	2.545093
37	Cmp_crwdsrcng	2.483072

7.

5. **Results**: AUC:

calculating AUC for X_test data ¶

```
2]: from sklearn.metrics import roc_auc_score

    # auc scores
    auc_score1 = roc_auc_score(Y_test, prid)
    auc_score1

2]: 0.9545454545454546
```