

Project 1 Group 12 - Mitali Bharali

Project description:

- Please read the Data Set Information section to learn about this dataset.
- Data description is also provided for thi dataset.
- Read data into Jupyter notebook, use pandas to import data into a data frame
- Preprocess data: Explore data, check for missing data and apply data scaling. Justify the type of scaling used.

Regression Task:

- Apply all the regression models you've learned so far. If your model has a scaling parameter(s) use Grid Search to find the best scaling parameter. Use plots and graphs to help you get a better glimpse of the results.
- Then use cross validation to find average training and testing score.
- Your submission should have at least the following regression models: KNN repressor, linear regression, Ridge, Lasso, polynomial regression, SVM both simple and with kernels.
- Finally find the best regressor for this dataset and train your model on the entire dataset using the best parameters and predict buzz for the test_set.

Classification task:

- Decide aboute a good evaluation strategy and justify your choice.
- Find best parameters for following classification models: KNN classifcation, Logistic Regression, Linear Supprt Vector Machine, Kerenilzed Support Vector Machine, Decision Tree.
- Which model gives the best results?

Deliverables:

- Submit IPython notebook. Use markdown to provide an inline comments for this project.
- Submit only one notebook. Before submitting, make sure everything runs as expected. To check that, restart the kernel (in the menubar, select Kernel > Restart) and then run all cells (in the menubar, select Cell > Run All).
- Visualization encouraged.

Questions regarding project:

- Post your queries related to project on discussion board on e-learning. There is high possibility that your classmate has also faced the same problem and knows the solution. This is an effort to encourage collaborative learning and also making all the information available to everyone. We will also answer queries there. We will not be answering any project related queries through mail.

Data Set Information:

This dataset is taken from a research explained here.

The goal of the research is to help the auditors by building a classification model that can predict the fraudulent firm on the basis the present and historical risk factors. The information about the sectors and the counts of firms are listed respectively as Irrigation (114), Public Health (77), Buildings and Roads (82), Forest (70), Corporate (47), Animal Husbandry (95), Communication (1), Electrical (4), Land (5), Science and Technology (3), Tourism (1), Fisheries (41), Industries (37), Agriculture (200).

There are two csv files to present data. Please merge these two datasets into one dataframe. All the steps should be done in Python. Please don't make any changes in csv files. Consider `Audit_Risk` as target columns for regression tasks, and `Risk` as the target column for classification tasks.

Attribute Information:

Many risk factors are examined from various areas like past records of audit office, audit-paras, environmental conditions reports, firm reputation summary, on-going issues report, profit-value records, loss-value records, follow-up reports etc. After in-depth interview with the auditors, important risk factors are evaluated and their probability of existence is calculated from the present and past records.

Relevant Papers:

Hooda, Nishtha, Seema Bawa, and Prashant Singh Rana. 'Fraudulent Firm Classification: A Case Study of an External Audit.' *Applied Artificial Intelligence* 32.1 (2018): 48-64.

Project Brief:

We went through the basic data processing for both the datasets, audit risk and trail. We found several columns derived from others, like Risk A,B,C.. are derivatives of Audit Risk column, Score_A, Score_B are same as SCORE_A and SCORE_B, just scaled differently. We also found duplicate columns in both Audit_Risk and Trail. Since there isn't any supporting data description, we took this purely in terms of Data Mining approach, doing as per what data tells us from the basic pre processing. The features chosen for the regression model was the main contributing factor alongside the removal of outliers in the model. The features were chosen by looking at the correlation heat map as well reference to the research paper.

The flow/order of work is described as below:

- Data Preprocessing - checking for na values, replacing missing values, incorrect values, dropping repeated columns.
- Merge - we used join using the common columns as index for the merge.
- Visualization - heat map to look for correl columns and box plot to identify outliers
- Regression and Classification tasks:
 - MODEL RUN
 - GRID SEARCH if required
 - Cross Validation- naive and Kfold
 - Model Evaluation for Classification- Precision, Recall, F1 score and Support; Accuracy for Regression

Note: Based on the reference from the research paper provided, we found that the Risk column calculated in the Audit_Risk table has been basically derived from the Audit Risk column in that table. That was removed as it was causing a lot of duplicate as well as null values while merging the Audit_Risk as well as the Trial table.

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

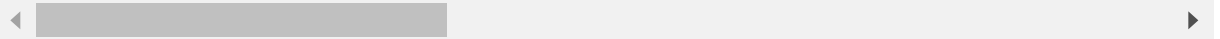
```
In [2]: audit_risk = pd.read_csv('audit_risk.csv')
trial = pd.read_csv('trial.csv')
```

In [3]: `audit_risk.describe()`

Out[3]:

	Sector_score	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B
count	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000
mean	20.184536	2.450194	0.351289	1.351029	10.799988	0.313144	6.334008
std	24.319017	5.678870	0.174055	3.440447	50.083624	0.169804	30.072845
min	1.850000	0.000000	0.200000	0.000000	0.000000	0.200000	0.000000
25%	2.370000	0.210000	0.200000	0.042000	0.000000	0.200000	0.000000
50%	3.890000	0.875000	0.200000	0.175000	0.405000	0.200000	0.081000
75%	55.570000	2.480000	0.600000	1.488000	4.160000	0.400000	1.840500
max	59.850000	85.000000	0.600000	51.000000	1264.630000	0.600000	758.778000

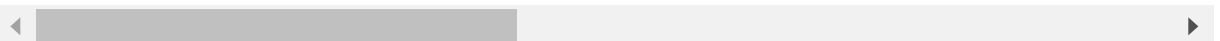
8 rows × 26 columns



In [4]: `trial.describe()`

Out[4]:

	Sector_score	PARA_A	SCORE_A	PARA_B	SCORE_B	TOTAL	numbers
count	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000	776.000000
mean	20.184536	2.450194	3.512887	10.799988	3.131443	13.218481	5.067655
std	24.319017	5.678870	1.740549	50.083624	1.698042	51.312829	0.264449
min	1.850000	0.000000	2.000000	0.000000	2.000000	0.000000	5.000000
25%	2.370000	0.210000	2.000000	0.000000	2.000000	0.537500	5.000000
50%	3.890000	0.875000	2.000000	0.405000	2.000000	1.370000	5.000000
75%	55.570000	2.480000	6.000000	4.160000	4.000000	7.707500	5.000000
max	59.850000	85.000000	6.000000	1264.630000	6.000000	1268.910000	9.000000



```
In [5]: audit_risk.isna().sum()
```

```
Out[5]: Sector_score      0
LOCATION_ID      0
PARA_A         0
Score_A        0
Risk_A         0
PARA_B         0
Score_B        0
Risk_B         0
TOTAL          0
numbers        0
Score_B.1      0
Risk_C         0
Money_Value    1
Score_MV       0
Risk_D         0
District_Loss  0
PROB           0
Risk_E         0
History        0
Prob           0
Risk_F         0
Score          0
Inherent_Risk  0
CONTROL_RISK   0
Detection_Risk 0
Audit_Risk     0
Risk           0
dtype: int64
```

```
In [6]: trial.isna().sum()
```

```
Out[6]: Sector_score      0
LOCATION_ID      0
PARA_A         0
SCORE_A        0
PARA_B         0
SCORE_B        0
TOTAL          0
numbers        0
Marks          0
Money_Value    1
MONEY_Marks    0
District       0
Loss           0
LOSS_SCORE     0
History        0
History_score  0
Score          0
Risk           0
dtype: int64
```

```
In [7]: nan_rows = audit_risk[audit_risk['Money_Value'].isnull()]
nan_rows
```

Out[7]:

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TOTAL
642	55.57	4	0.23	0.2	0.046	0.0	0.2	0.0	0.2

1 rows × 27 columns



```
In [8]: nan_rows_trial = trial[trial['Money_Value'].isnull()]
nan_rows_trial
```

Out[8]:

	Sector_score	LOCATION_ID	PARA_A	SCORE_A	PARA_B	SCORE_B	TOTAL	numbers	M
642	55.57	4	0.23	2	0.0	2	0.23	5.0	



```
In [9]: audit_risk.dtypes
```

```
Out[9]: Sector_score      float64
LOCATION_ID      object
PARA_A         float64
Score_A        float64
Risk_A         float64
PARA_B         float64
Score_B        float64
Risk_B         float64
TOTAL          float64
numbers        float64
Score_B.1      float64
Risk_C         float64
Money_Value    float64
Score_MV       float64
Risk_D         float64
District_Loss  int64
PROB           float64
Risk_E         float64
History        int64
Prob           float64
Risk_F         float64
Score          float64
Inherent_Risk  float64
CONTROL_RISK   float64
Detection_Risk float64
Audit_Risk     float64
Risk           int64
dtype: object
```

```
In [10]: audit_risk.fillna(audit_risk.median(), inplace=True)
```

```
In [11]: audit_risk.isna().sum()
```

```
Out[11]: Sector_score      0
LOCATION_ID      0
PARA_A         0
Score_A        0
Risk_A         0
PARA_B         0
Score_B        0
Risk_B         0
TOTAL          0
numbers        0
Score_B.1      0
Risk_C         0
Money_Value    0
Score_MV       0
Risk_D         0
District_Loss  0
PROB           0
Risk_E         0
History        0
Prob           0
Risk_F         0
Score          0
Inherent_Risk  0
CONTROL_RISK   0
Detection_Risk 0
Audit_Risk     0
Risk           0
dtype: int64
```

```
In [12]: trial.fillna(trial.median(), inplace=True)
```

```
In [13]: trial.isna().sum()
```

```
Out[13]: Sector_score      0
LOCATION_ID      0
PARA_A         0
SCORE_A        0
PARA_B         0
SCORE_B        0
TOTAL          0
numbers        0
Marks          0
Money_Value    0
MONEY_Marks    0
District       0
Loss           0
LOSS_SCORE     0
History        0
History_score  0
Score          0
Risk           0
dtype: int64
```

```
In [14]: audit_risk.info()  
trial.info()
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 776 entries, 0 to 775
Data columns (total 27 columns):
Sector_score      776 non-null float64
LOCATION_ID        776 non-null object
PARA_A           776 non-null float64
Score_A          776 non-null float64
Risk_A           776 non-null float64
PARA_B           776 non-null float64
Score_B          776 non-null float64
Risk_B           776 non-null float64
TOTAL            776 non-null float64
numbers          776 non-null float64
Score_B.1        776 non-null float64
Risk_C           776 non-null float64
Money_Value      776 non-null float64
Score_MV         776 non-null float64
Risk_D           776 non-null float64
District_Loss    776 non-null int64
PROB             776 non-null float64
Risk_E           776 non-null float64
History          776 non-null int64
Prob            776 non-null float64
Risk_F           776 non-null float64
Score            776 non-null float64
Inherent_Risk    776 non-null float64
CONTROL_RISK     776 non-null float64
Detection_Risk   776 non-null float64
Audit_Risk       776 non-null float64
Risk             776 non-null int64
dtypes: float64(23), int64(3), object(1)
memory usage: 163.8+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 776 entries, 0 to 775
Data columns (total 18 columns):
Sector_score      776 non-null float64
LOCATION_ID        776 non-null object
PARA_A           776 non-null float64
SCORE_A          776 non-null int64
PARA_B           776 non-null float64
SCORE_B          776 non-null int64
TOTAL            776 non-null float64
numbers          776 non-null float64
Marks            776 non-null int64
Money_Value      776 non-null float64
MONEY_Marks      776 non-null int64
District         776 non-null int64
Loss             776 non-null int64
LOSS_SCORE       776 non-null int64
History          776 non-null int64
History_score     776 non-null int64
Score            776 non-null float64
Risk             776 non-null int64
dtypes: float64(7), int64(10), object(1)
memory usage: 109.2+ KB
```

```
In [15]: audit_risk.describe(include = 'O')
```

Out[15]:

LOCATION_ID	
count	776
unique	45
top	8
freq	76

```
In [16]: audit_risk.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 776 entries, 0 to 775
Data columns (total 27 columns):
Sector_score      776 non-null float64
LOCATION_ID        776 non-null object
PARA_A           776 non-null float64
Score_A          776 non-null float64
Risk_A           776 non-null float64
PARA_B           776 non-null float64
Score_B          776 non-null float64
Risk_B           776 non-null float64
TOTAL            776 non-null float64
numbers          776 non-null float64
Score_B.1        776 non-null float64
Risk_C           776 non-null float64
Money_Value      776 non-null float64
Score_MV         776 non-null float64
Risk_D           776 non-null float64
District_Loss    776 non-null int64
PROB             776 non-null float64
Risk_E           776 non-null float64
History          776 non-null int64
Prob            776 non-null float64
Risk_F           776 non-null float64
Score            776 non-null float64
Inherent_Risk    776 non-null float64
CONTROL_RISK     776 non-null float64
Detection_Risk   776 non-null float64
Audit_Risk       776 non-null float64
Risk             776 non-null int64
dtypes: float64(23), int64(3), object(1)
memory usage: 163.8+ KB
```

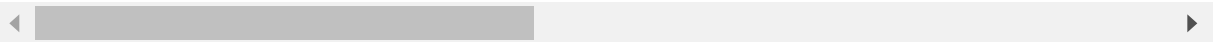
```
In [17]: audit_risk=audit_risk.drop(['Risk'],axis=1)
         audit_risk
```

Out[17]:

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TO
0	3.89	23	4.18	0.6	2.508	2.5000	0.2	0.50000	6.6
1	3.89	6	0.00	0.2	0.000	4.8300	0.2	0.96600	4.8
2	3.89	6	0.51	0.2	0.102	0.2300	0.2	0.04600	0.7
3	3.89	6	0.00	0.2	0.000	10.8000	0.6	6.48000	10.8
4	3.89	6	0.00	0.2	0.000	0.0800	0.2	0.01600	0.0
5	3.89	6	0.00	0.2	0.000	0.8300	0.2	0.16600	0.8
6	3.89	7	1.10	0.4	0.440	7.4100	0.4	2.96400	8.5
7	3.89	8	8.50	0.6	5.100	12.0300	0.6	7.21800	20.5
8	3.89	8	8.40	0.6	5.040	11.0500	0.6	6.63000	19.4
9	3.89	8	3.98	0.6	2.388	0.9900	0.2	0.19800	4.9
10	3.89	8	5.43	0.6	3.258	10.7700	0.6	6.46200	16.2
11	3.89	8	15.38	0.6	9.228	40.1400	0.6	24.08400	55.5
12	3.89	8	5.47	0.6	3.282	7.6300	0.4	3.05200	13.1
13	3.89	8	1.09	0.4	0.436	0.3500	0.2	0.07000	1.4
14	3.89	8	0.00	0.2	0.000	0.8400	0.2	0.16800	0.8
15	3.89	13	1.95	0.4	0.780	9.0100	0.4	3.60400	10.9
16	3.89	37	8.54	0.6	5.124	31.6300	0.6	18.97800	40.1
17	3.89	37	4.18	0.6	2.508	4.8300	0.2	0.96600	9.0
18	3.89	37	1.81	0.4	0.724	1.0300	0.2	0.20600	2.8
19	3.89	37	4.86	0.6	2.916	46.7800	0.6	28.06800	51.6
20	3.89	24	6.26	0.6	3.756	14.1000	0.6	8.46000	20.3
21	3.89	3	0.02	0.2	0.004	5.9400	0.4	2.37600	5.9
22	3.89	3	5.31	0.6	3.186	22.7900	0.6	13.67400	28.1
23	3.89	3	0.94	0.2	0.188	0.0100	0.2	0.00200	0.9
24	3.89	4	5.78	0.6	3.468	57.9200	0.6	34.75200	63.7
25	3.89	4	7.42	0.6	4.452	2.2400	0.2	0.44800	9.6
26	3.89	4	0.00	0.2	0.000	1.1000	0.2	0.22000	1.1
27	3.89	14	6.85	0.6	4.110	31.7600	0.6	19.05600	38.6
28	3.89	14	0.00	0.2	0.000	1.0300	0.2	0.20600	1.0
29	3.89	37	0.00	0.2	0.000	0.7500	0.2	0.15000	0.7
...
746	55.57	13	0.25	0.2	0.050	0.0017	0.2	0.00034	0.2
747	55.57	13	0.31	0.2	0.062	0.0015	0.2	0.00030	0.3
748	55.57	13	0.00	0.2	0.000	0.0000	0.2	0.00000	0.0
749	55.57	13	0.84	0.2	0.168	0.0000	0.2	0.00000	0.8

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TO
750	55.57	13	1.09	0.4	0.436	0.0000	0.2	0.00000	1.0
751	55.57	13	1.29	0.4	0.516	0.0000	0.2	0.00000	1.2
752	55.57	13	0.51	0.2	0.102	0.3700	0.2	0.07400	0.8
753	55.57	21	0.09	0.2	0.018	0.0000	0.2	0.00000	0.0
754	55.57	18	0.39	0.2	0.078	0.9100	0.2	0.18200	1.3
755	55.57	21	1.07	0.4	0.428	0.0000	0.2	0.00000	1.0
756	55.57	25	0.00	0.2	0.000	0.0000	0.2	0.00000	0.0
757	55.57	32	0.50	0.2	0.100	2.9700	0.6	1.78200	3.4
758	55.57	22	0.49	0.2	0.098	0.5500	0.2	0.11000	1.0
759	55.57	14	0.84	0.2	0.168	0.6500	0.2	0.13000	1.4
760	55.57	12	0.90	0.2	0.180	1.1100	0.4	0.44400	2.0
761	55.57	12	0.00	0.2	0.000	0.0000	0.2	0.00000	0.0
762	55.57	14	0.59	0.2	0.118	0.0000	0.2	0.00000	0.5
763	55.57	36	0.02	0.2	0.004	0.0000	0.2	0.00000	0.0
764	55.57	14	1.48	0.4	0.592	4.4800	0.6	2.68800	5.9
765	55.57	22	0.00	0.2	0.000	3.3000	0.6	1.98000	3.3
766	55.57	8	0.80	0.2	0.160	0.5700	0.2	0.11400	1.3
767	55.57	18	0.36	0.2	0.072	0.5400	0.2	0.10800	0.9
768	55.57	9	0.44	0.2	0.088	0.5300	0.2	0.10600	0.9
769	55.57	16	0.51	0.2	0.102	0.5000	0.2	0.10000	1.0
770	55.57	18	0.75	0.2	0.150	0.4500	0.2	0.09000	1.2
771	55.57	9	0.49	0.2	0.098	0.4000	0.2	0.08000	0.8
772	55.57	16	0.47	0.2	0.094	0.3700	0.2	0.07400	0.8
773	55.57	14	0.24	0.2	0.048	0.0400	0.2	0.00800	0.2
774	55.57	18	0.20	0.2	0.040	0.0000	0.2	0.00000	0.2
775	55.57	15	0.00	0.2	0.000	0.0000	0.2	0.00000	0.0

776 rows × 26 columns



```
In [18]: audit_risk = audit_risk.drop_duplicates(keep='first')
         audit_risk.shape
```

Out[18]: (763, 26)

```
In [19]: trial = trial.drop_duplicates(keep='first')
         trial.shape
```

Out[19]: (763, 18)

```
In [20]: result = audit_risk.join(trial.set_index(['Sector_score', 'LOCATION_ID', 'PARA_A', 'PARA_B', 'TOTAL', 'numbers', 'Money_Value', 'History', 'Score']), on= ['Sector_score', 'LOCATION_ID', 'PARA_A', 'PARA_B', 'TOTAL', 'numbers', 'Money_Value', 'History', 'Score'])
```

```
In [21]: result=result.drop(['SCORE_A', 'SCORE_B'],axis=1)
```

```
In [22]: result.isna().sum()
```

```
Out[22]: Sector_score      0
LOCATION_ID      0
PARA_A          0
Score_A         0
Risk_A          0
PARA_B          0
Score_B         0
Risk_B          0
TOTAL           0
numbers         0
Score_B.1       0
Risk_C          0
Money_Value     0
Score_MV        0
Risk_D          0
District_Loss   0
PROB            0
Risk_E          0
History         0
Prob           0
Risk_F          0
Score           0
Inherent_Risk   0
CONTROL_RISK    0
Detection_Risk  0
Audit_Risk      0
Marks           0
MONEY_Marks     0
District        0
Loss            0
LOSS_SCORE      0
History_score   0
Risk            0
dtype: int64
```

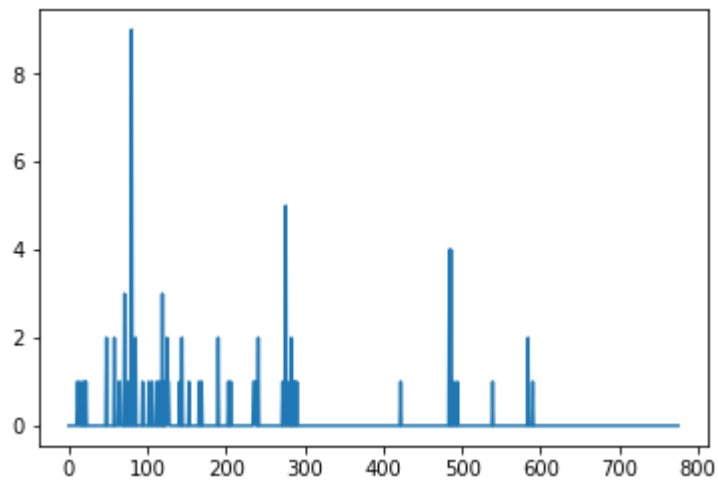
In [23]: result.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 763 entries, 0 to 775
Data columns (total 33 columns):
Sector_score      763 non-null float64
LOCATION_ID        763 non-null object
PARA_A            763 non-null float64
Score_A           763 non-null float64
Risk_A            763 non-null float64
PARA_B            763 non-null float64
Score_B           763 non-null float64
Risk_B            763 non-null float64
TOTAL             763 non-null float64
numbers           763 non-null float64
Score_B.1         763 non-null float64
Risk_C            763 non-null float64
Money_Value       763 non-null float64
Score_MV          763 non-null float64
Risk_D            763 non-null float64
District_Loss     763 non-null int64
PROB              763 non-null float64
Risk_E            763 non-null float64
History           763 non-null int64
Prob              763 non-null float64
Risk_F            763 non-null float64
Score             763 non-null float64
Inherent_Risk     763 non-null float64
CONTROL_RISK      763 non-null float64
Detection_Risk    763 non-null float64
Audit_Risk        763 non-null float64
Marks             763 non-null int64
MONEY_Marks       763 non-null int64
District          763 non-null int64
Loss              763 non-null int64
LOSS_SCORE        763 non-null int64
History_score     763 non-null int64
Risk              763 non-null int64
dtypes: float64(23), int64(9), object(1)
memory usage: 202.7+ KB
```

In [24]: result = result.dropna()

```
In [25]: plt.plot(audit_risk['History'])
```

```
Out[25]: [<matplotlib.lines.Line2D at 0x21adc5d4c50>]
```



```
In [26]: result.shape
```

```
Out[26]: (763, 33)
```



```
In [27]: dup = result.duplicated(subset=None, keep = 'first')  
dup
```

```
Out[27]: 0      False
          1      False
          2      False
          3      False
          4      False
          5      False
          6      False
          7      False
          8      False
          9      False
         10      False
         11      False
         12      False
         13      False
         14      False
         15      False
         16      False
         17      False
         18      False
         19      False
         20      False
         21      False
         22      False
         23      False
         24      False
         25      False
         26      False
         27      False
         28      False
         29      False
          ...
        744      False
        745      False
        746      False
        747      False
        748      False
        749      False
        750      False
        751      False
        752      False
        753      False
        754      False
        755      False
        756      False
        757      False
        758      False
        759      False
        760      False
        762      False
        763      False
        764      False
        765      False
        766      False
        767      False
        768      False
        769      False
        770      False
```

```
772 False
773 False
774 False
775 False
Length: 763, dtype: bool
```

```
In [28]: dup.sum()
```

```
Out[28]: 0
```

```
In [29]: result=result.drop_duplicates(keep='first')
```

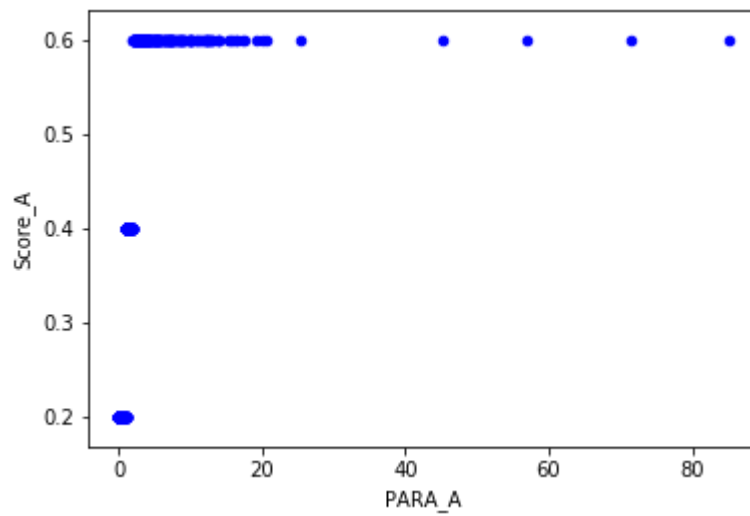
```
In [30]: result.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 763 entries, 0 to 775
Data columns (total 33 columns):
Sector_score      763 non-null float64
LOCATION_ID        763 non-null object
PARA_A           763 non-null float64
Score_A          763 non-null float64
Risk_A           763 non-null float64
PARA_B           763 non-null float64
Score_B          763 non-null float64
Risk_B           763 non-null float64
TOTAL            763 non-null float64
numbers          763 non-null float64
Score_B.1        763 non-null float64
Risk_C           763 non-null float64
Money_Value      763 non-null float64
Score_MV         763 non-null float64
Risk_D           763 non-null float64
District_Loss    763 non-null int64
PROB             763 non-null float64
Risk_E           763 non-null float64
History          763 non-null int64
Prob            763 non-null float64
Risk_F           763 non-null float64
Score            763 non-null float64
Inherent_Risk    763 non-null float64
CONTROL_RISK     763 non-null float64
Detection_Risk   763 non-null float64
Audit_Risk       763 non-null float64
Marks            763 non-null int64
MONEY_Marks      763 non-null int64
District         763 non-null int64
Loss             763 non-null int64
LOSS_SCORE       763 non-null int64
History_score    763 non-null int64
Risk             763 non-null int64
dtypes: float64(23), int64(9), object(1)
memory usage: 202.7+ KB
```

```
In [ ]:
```

```
In [31]: result.plot(kind='scatter',x='PARAM_A',y='Score_A',color='blue')
```

```
Out[31]: <matplotlib.axes._subplots.AxesSubplot at 0x21adc62c1d0>
```



```
In [ ]:
```

```
In [32]: result.LOCATION_ID.replace(['LOHARU', 'NUH', 'SAFIDON'], [10, 26, 45], inplace
        =True)
        result.dtypes
```

```
Out[32]: Sector_score      float64
LOCATION_ID      object
PARA_A         float64
Score_A        float64
Risk_A         float64
PARA_B         float64
Score_B        float64
Risk_B         float64
TOTAL          float64
numbers        float64
Score_B.1      float64
Risk_C         float64
Money_Value    float64
Score_MV       float64
Risk_D         float64
District_Loss  int64
PROB           float64
Risk_E         float64
History        int64
Prob           float64
Risk_F         float64
Score          float64
Inherent_Risk  float64
CONTROL_RISK   float64
Detection_Risk float64
Audit_Risk     float64
Marks          int64
MONEY_Marks    int64
District       int64
Loss           int64
LOSS_SCORE     int64
History_score  int64
Risk           int64
dtype: object
```

```
In [33]: result["LOCATION_ID"]=result["LOCATION_ID"].astype(int)
result.dtypes
```

```
Out[33]: Sector_score      float64
LOCATION_ID      int32
PARA_A         float64
Score_A        float64
Risk_A         float64
PARA_B         float64
Score_B        float64
Risk_B         float64
TOTAL          float64
numbers        float64
Score_B.1      float64
Risk_C         float64
Money_Value    float64
Score_MV       float64
Risk_D         float64
District_Loss  int64
PROB           float64
Risk_E         float64
History        int64
Prob           float64
Risk_F         float64
Score          float64
Inherent_Risk  float64
CONTROL_RISK   float64
Detection_Risk float64
Audit_Risk     float64
Marks          int64
MONEY_Marks    int64
District       int64
Loss           int64
LOSS_SCORE     int64
History_score  int64
Risk           int64
dtype: object
```

```
In [34]: def plot_corr(result,size=10):
        '''Function plots a graphical correlation matrix for each pair of columns
        in the dataframe.

        Input:
            df: pandas DataFrame
            size: vertical and horizontal size of the plot'''

        corr = result.corr()
        fig, ax = plt.subplots(figsize=(size, size))
        ax.matshow(corr)
        plt.xticks(range(len(corr.columns)), corr.columns);
        plt.yticks(range(len(corr.columns)), corr.columns);
```

```
In [35]: rc = result.corr()  
         print(rc)
```

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	\
Sector_score	1.000000	-0.054881	-0.212488	-0.417980	-0.215317	
LOCATION_ID	-0.054881	1.000000	-0.008273	0.074465	-0.006064	
PARA_A	-0.212488	-0.008273	1.000000	0.494335	0.999277	
Score_A	-0.417980	0.074465	0.494335	1.000000	0.502990	
Risk_A	-0.215317	-0.006064	0.999277	0.502990	1.000000	
PARA_B	-0.130719	0.005488	0.160693	0.247950	0.164133	
Score_B	-0.210417	0.121297	0.355618	0.568595	0.360367	
Risk_B	-0.128688	0.005144	0.160618	0.246756	0.164016	
TOTAL	-0.150939	0.005253	0.265743	0.295824	0.269016	
numbers	-0.149438	0.005455	0.132210	0.240587	0.134158	
Score_B.1	-0.167585	-0.018951	0.138827	0.272179	0.141426	
Risk_C	-0.163772	-0.016387	0.138702	0.263648	0.141093	
Money_Value	-0.115674	-0.067059	0.448695	0.204516	0.448094	
Score_MV	-0.313937	0.103871	0.283084	0.474807	0.288129	
Risk_D	-0.114090	-0.067431	0.447912	0.201991	0.447281	
District_Loss	-0.109003	-0.106627	0.127700	0.087413	0.127310	
PROB	-0.084987	-0.004209	0.042588	0.091817	0.042701	
Risk_E	-0.128644	-0.098154	0.118286	0.100879	0.118056	
History	-0.112895	-0.082191	0.117074	0.176214	0.120019	
Prob	-0.134424	-0.055019	0.171202	0.263507	0.174835	
Risk_F	-0.101534	-0.089318	0.102896	0.149330	0.105559	
Score	-0.329635	0.083136	0.423919	0.717350	0.430160	
Inherent_Risk	-0.170498	-0.047922	0.480815	0.318829	0.482307	
CONTROL_RISK	-0.153803	-0.125519	0.147995	0.168668	0.149681	
Detection_Risk	NaN	NaN	NaN	NaN	NaN	
Audit_Risk	-0.090077	-0.086868	0.218880	0.200587	0.220754	
Marks	-0.167585	-0.018951	0.138827	0.272179	0.141426	
MONEY_Marks	-0.313937	0.103871	0.283084	0.474807	0.288129	
District	-0.109003	-0.106627	0.127700	0.087413	0.127310	
Loss	-0.081408	0.005591	0.044306	0.091572	0.044863	
LOSS_SCORE	-0.084987	-0.004209	0.042588	0.091817	0.042701	
History_score	-0.134424	-0.055019	0.171202	0.263507	0.174835	
Risk	-0.364573	0.038772	0.288588	0.668401	0.286694	

	PARA_B	Score_B	Risk_B	TOTAL	numbers	...	\
Sector_score	-0.130719	-0.210417	-0.128688	-0.150939	-0.149438	...	
LOCATION_ID	0.005488	0.121297	0.005144	0.005253	0.005455	...	
PARA_A	0.160693	0.355618	0.160618	0.265743	0.132210	...	
Score_A	0.247950	0.568595	0.246756	0.295824	0.240587	...	
Risk_A	0.164133	0.360367	0.164016	0.269016	0.134158	...	
PARA_B	1.000000	0.346575	0.999936	0.994109	0.209116	...	
Score_B	0.346575	1.000000	0.347208	0.378068	0.275868	...	
Risk_B	0.999936	0.347208	1.000000	0.994035	0.208873	...	
TOTAL	0.994109	0.378068	0.994035	1.000000	0.216765	...	
numbers	0.209116	0.275868	0.208873	0.216765	1.000000	...	
Score_B.1	0.230030	0.311858	0.229745	0.237256	0.908033	...	
Risk_C	0.222251	0.301124	0.221957	0.229741	0.955255	...	
Money_Value	0.124403	0.203696	0.124452	0.167957	0.185493	...	
Score_MV	0.313450	0.564819	0.312027	0.336096	0.445663	...	
Risk_D	0.124006	0.200657	0.124066	0.167483	0.185837	...	
District_Loss	0.082933	-0.006549	0.083129	0.093308	0.125199	...	
PROB	0.041972	0.091058	0.042309	0.045786	0.035115	...	
Risk_E	0.079085	0.013186	0.079437	0.088824	0.136833	...	
History	0.203008	0.199429	0.202260	0.210080	0.201648	...	
Prob	0.315925	0.308031	0.315773	0.325670	0.208747	...	
Risk_F	0.195534	0.169967	0.194788	0.201559	0.201786	...	

Score	0.396568	0.900613	0.396414	0.432972	0.502386	...
Inherent_Risk	0.654114	0.363949	0.654178	0.689380	0.270176	...
CONTROL_RISK	0.186326	0.125771	0.186041	0.196824	0.228255	...
Detection_Risk	NaN	NaN	NaN	NaN	NaN	...
Audit_Risk	0.887734	0.206814	0.887508	0.888060	0.220874	...
Marks	0.230030	0.311858	0.229745	0.237256	0.908033	...
MONEY_Marks	0.313450	0.564819	0.312027	0.336096	0.445663	...
District	0.082933	-0.006549	0.083129	0.093308	0.125199	...
Loss	0.044251	0.097432	0.044558	0.048197	-0.002188	...
LOSS_SCORE	0.041972	0.091058	0.042309	0.045786	0.035115	...
History_score	0.315925	0.308031	0.315773	0.325670	0.208747	...
Risk	0.160900	0.511017	0.159674	0.188507	0.195569	...

	CONTROL_RISK	Detection_Risk	Audit_Risk	Marks	\
Sector_score	-0.153803	NaN	-0.090077	-0.167585	
LOCATION_ID	-0.125519	NaN	-0.086868	-0.018951	
PARA_A	0.147995	NaN	0.218880	0.138827	
Score_A	0.168668	NaN	0.200587	0.272179	
Risk_A	0.149681	NaN	0.220754	0.141426	
PARA_B	0.186326	NaN	0.887734	0.230030	
Score_B	0.125771	NaN	0.206814	0.311858	
Risk_B	0.186041	NaN	0.887508	0.229745	
TOTAL	0.196824	NaN	0.888060	0.237256	
numbers	0.228255	NaN	0.220874	0.908033	
Score_B.1	0.256114	NaN	0.259112	1.000000	
Risk_C	0.251150	NaN	0.249409	0.990399	
Money_Value	0.068913	NaN	0.333663	0.219590	
Score_MV	0.216156	NaN	0.290894	0.506159	
Risk_D	0.068898	NaN	0.333773	0.219843	
District_Loss	0.643255	NaN	0.200204	0.150658	
PROB	0.338443	NaN	0.073371	0.036104	
Risk_E	0.728427	NaN	0.203344	0.157480	
History	0.755577	NaN	0.329378	0.225430	
Prob	0.642710	NaN	0.430907	0.247494	
Risk_F	0.762193	NaN	0.327396	0.222685	
Score	0.355182	NaN	0.332500	0.565489	
Inherent_Risk	0.172515	NaN	0.750774	0.307511	
CONTROL_RISK	1.000000	NaN	0.358182	0.256114	
Detection_Risk	NaN	NaN	NaN	NaN	
Audit_Risk	0.358182	NaN	1.000000	0.259112	
Marks	0.256114	NaN	0.259112	1.000000	
MONEY_Marks	0.216156	NaN	0.290894	0.506159	
District	0.643255	NaN	0.200204	0.150658	
Loss	0.277699	NaN	0.048845	0.003955	
LOSS_SCORE	0.338443	NaN	0.073371	0.036104	
History_score	0.642710	NaN	0.430907	0.247494	
Risk	0.295708	NaN	0.135449	0.225623	

	MONEY_Marks	District	Loss	LOSS_SCORE	History_score	\
Sector_score	-0.313937	-0.109003	-0.081408	-0.084987	-0.134424	
LOCATION_ID	0.103871	-0.106627	0.005591	-0.004209	-0.055019	
PARA_A	0.283084	0.127700	0.044306	0.042588	0.171202	
Score_A	0.474807	0.087413	0.091572	0.091817	0.263507	
Risk_A	0.288129	0.127310	0.044863	0.042701	0.174835	
PARA_B	0.313450	0.082933	0.044251	0.041972	0.315925	
Score_B	0.564819	-0.006549	0.097432	0.091058	0.308031	
Risk_B	0.312027	0.083129	0.044558	0.042309	0.315773	

TOTAL	0.336096	0.093308	0.048197	0.045786	0.325670
numbers	0.445663	0.125199	-0.002188	0.035115	0.208747
Score_B.1	0.506159	0.150658	0.003955	0.036104	0.247494
Risk_C	0.491932	0.146522	0.000285	0.035448	0.236813
Money_Value	0.390606	0.027951	0.024007	0.031247	0.111452
Score_MV	1.000000	0.080284	0.117463	0.128705	0.333416
Risk_D	0.390261	0.028124	0.024081	0.031350	0.110715
District_Loss	0.080284	1.000000	0.036170	0.055227	0.083839
PROB	0.128705	0.055227	0.981502	1.000000	0.140883
RiSk_E	0.103296	0.912644	0.369421	0.407786	0.123525
History	0.245733	0.069317	0.060137	0.109571	0.859198
Prob	0.333416	0.083839	0.103904	0.140883	1.000000
Risk_F	0.215924	0.070687	0.053737	0.105594	0.815510
Score	0.756871	0.209063	0.155573	0.166452	0.442709
Inherent_Risk	0.481191	0.080846	0.047138	0.051913	0.272914
CONTROL_RISK	0.216156	0.643255	0.277699	0.338443	0.642710
Detection_Risk	NaN	NaN	NaN	NaN	NaN
Audit_Risk	0.290894	0.200204	0.048845	0.073371	0.430907
Marks	0.506159	0.150658	0.003955	0.036104	0.247494
MONEY_Marks	1.000000	0.080284	0.117463	0.128705	0.333416
District	0.080284	1.000000	0.036170	0.055227	0.083839
Loss	0.117463	0.036170	1.000000	0.981502	0.103904
LOSS_SCORE	0.128705	0.055227	0.981502	1.000000	0.140883
History_score	0.333416	0.083839	0.103904	0.140883	1.000000
Risk	0.436335	0.312959	0.122909	0.126025	0.188357

	Risk
Sector_score	-0.364573
LOCATION_ID	0.038772
PARA_A	0.288588
Score_A	0.668401
Risk_A	0.286694
PARA_B	0.160900
Score_B	0.511017
Risk_B	0.159674
TOTAL	0.188507
numbers	0.195569
Score_B.1	0.225623
Risk_C	0.217610
Money_Value	0.159010
Score_MV	0.436335
Risk_D	0.157010
District_Loss	0.312959
PROB	0.126025
RiSk_E	0.312168
History	0.150227
Prob	0.188357
Risk_F	0.133955
Score	0.627851
Inherent_Risk	0.225050
CONTROL_RISK	0.295708
Detection_Risk	NaN
Audit_Risk	0.135449
Marks	0.225623
MONEY_Marks	0.436335
District	0.312959
Loss	0.122909

```
LOSS_SCORE      0.126025
History_score    0.188357
Risk             1.000000
```

```
[33 rows x 33 columns]
```

```
In [36]: x = result[['Sector_score', 'PARA_A', 'PARA_B', 'numbers', 'Money_Value', 'District_Loss', 'History', 'Score', 'Audit_Risk', 'Loss']]
```

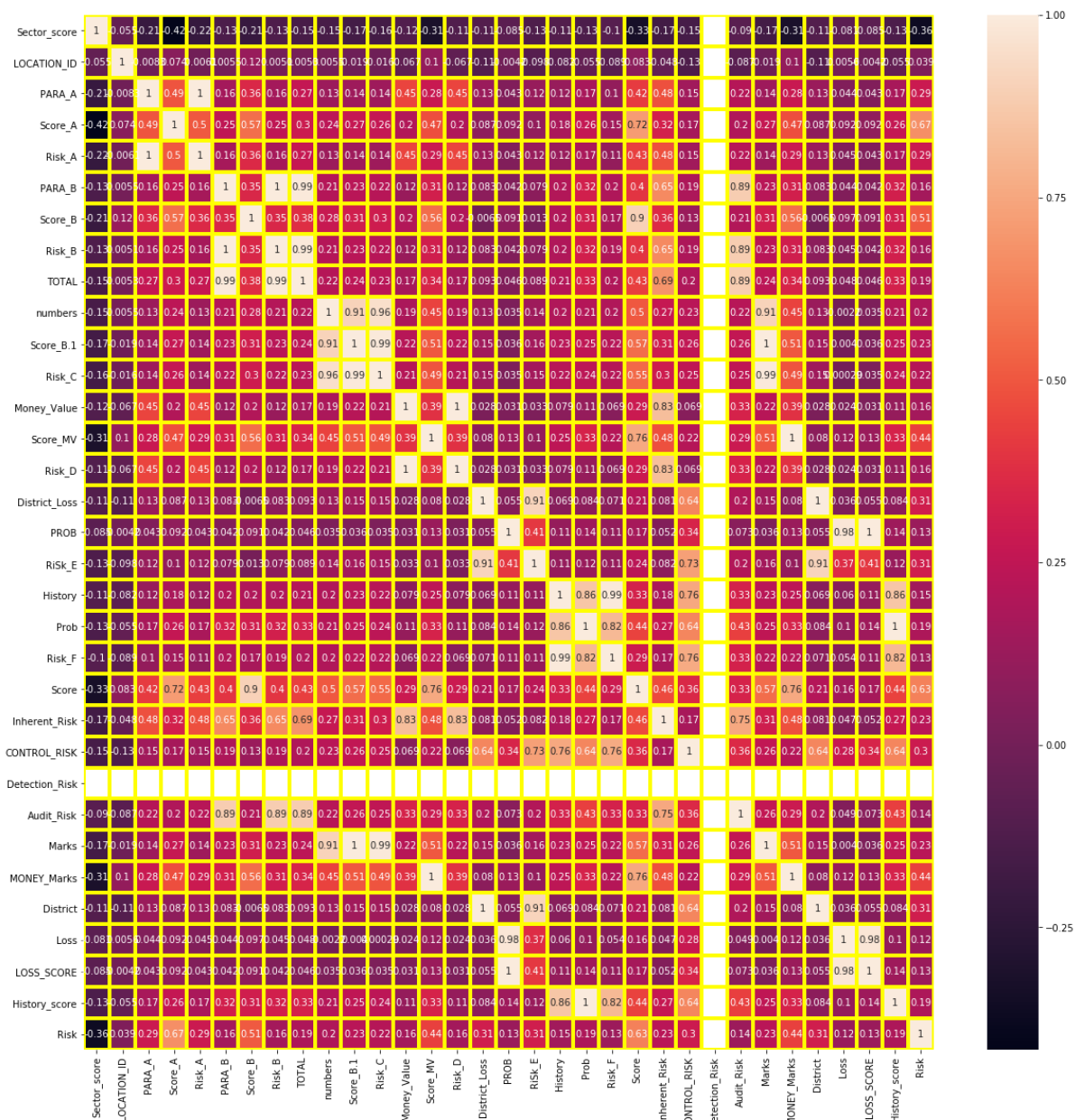
```
In [37]: #x = result[['PARA_A', 'numbers', 'History', 'PARA_B', 'Loss', 'Score', 'PROB', 'Money_Value', 'MONEY_Marks', 'District', 'TOTAL', 'LOCATION_ID']]
y = result[['Audit_Risk']]
```

VISUALIZATION OF DATA

```
In [38]: #CORREL MATRIX
import seaborn as sns
```

```
fig=plt.figure(figsize=(20,20))
sns.heatmap(rc,annot=True,linewidth='yellow',linewidths=3)
```

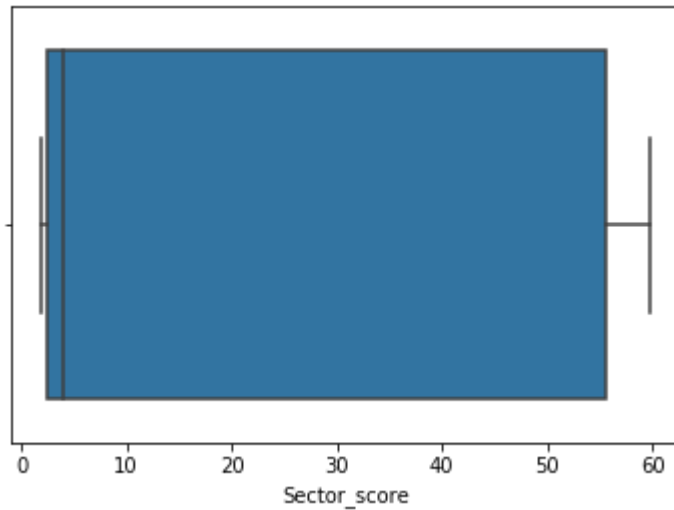
```
Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade255f60>
```



```
In [39]: x = result[['Sector_score', 'PARA_A', 'PARA_B', 'numbers', 'Money_Value', 'History',
, 'Score', 'District_Loss', 'Loss']]
```

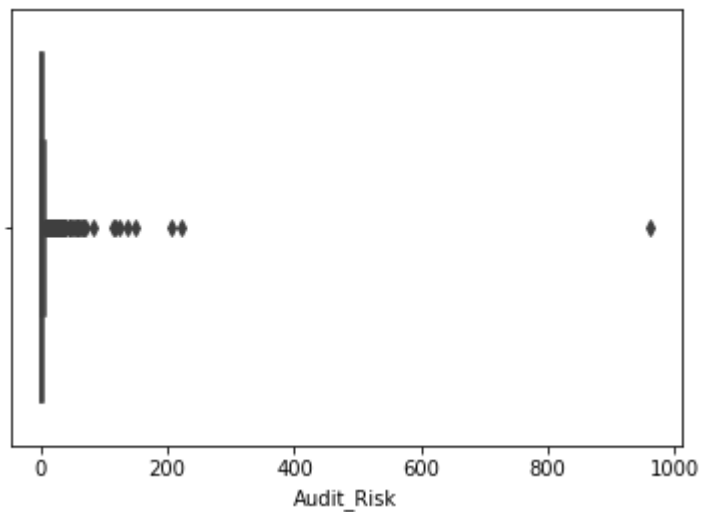
```
In [40]: import seaborn as sns  
sns.boxplot(x=result['Sector_score'])
```

Out[40]: <matplotlib.axes._subplots.AxesSubplot at 0x21adf1c3208>



```
In [41]: import seaborn as sns  
sns.boxplot(x=result['Audit_Risk'])
```

Out[41]: <matplotlib.axes._subplots.AxesSubplot at 0x21adf272240>



```
In [42]: result['Audit_Risk'].describe()
```

Out[42]:

count	763.000000
mean	7.284109
std	38.985618
min	0.280000
25%	0.318600
50%	0.590000
75%	3.452800
max	961.514400

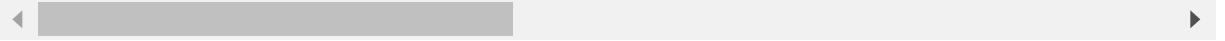
Name: Audit_Risk, dtype: float64

```
In [43]: result.loc[result['Audit_Risk'] == 961.514400]
```

```
Out[43]:
```

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TOT
241	2.72	2	4.28	0.6	2.568	1264.63	0.6	758.778	1268.

1 rows × 33 columns

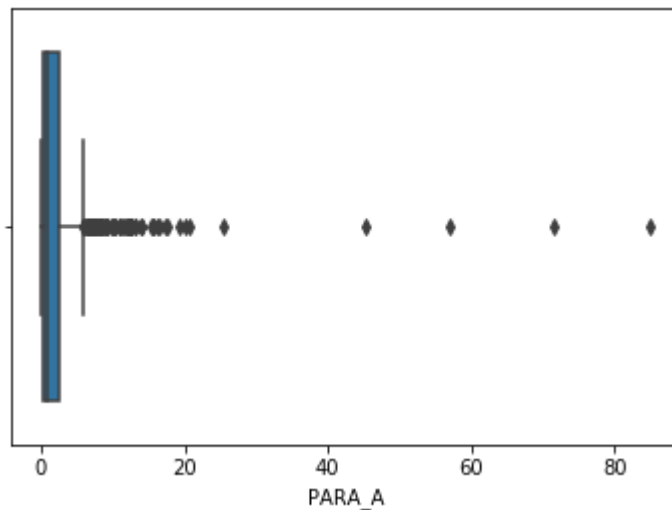


```
In [44]: i = result[result['Audit_Risk'] == 961.514400].index
```

```
In [45]: result = result.drop(i)
```

```
In [46]: #PARA_A
sns.boxplot(x=result['PARA_A'])
```

```
Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade54af98>
```



```
In [47]: result['PARA_A'].describe()
```

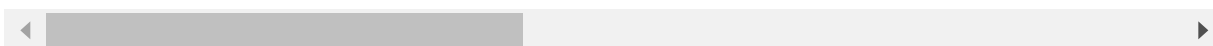
```
Out[47]: count      762.000000
mean         2.488951
std          5.721647
min          0.000000
25%          0.260000
50%          0.900000
75%          2.505000
max          85.000000
Name: PARA_A, dtype: float64
```

```
In [48]: result.loc[result['PARA_A'] > 40.000000]
```

```
Out[48]:
```

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TOTAL
81	3.89	2	71.48	0.6	42.888	25.63	0.6	15.378	97.1
342	1.99	2	57.03	0.6	34.218	134.33	0.6	80.598	191.3
467	1.85	19	45.23	0.6	27.138	36.18	0.6	21.708	81.4
509	21.61	9	85.00	0.6	51.000	1.06	0.4	0.424	86.0

4 rows × 33 columns

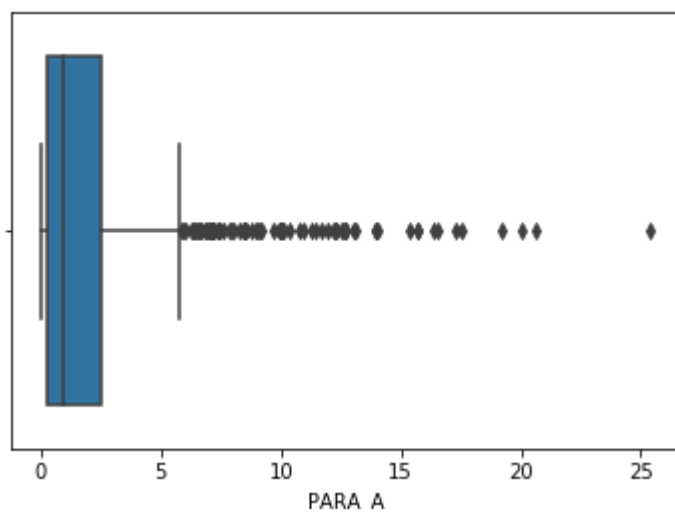


```
In [49]: i = result.loc[result['PARA_A'] > 40.000000].index
```

```
In [50]: result = result.drop(i)
```

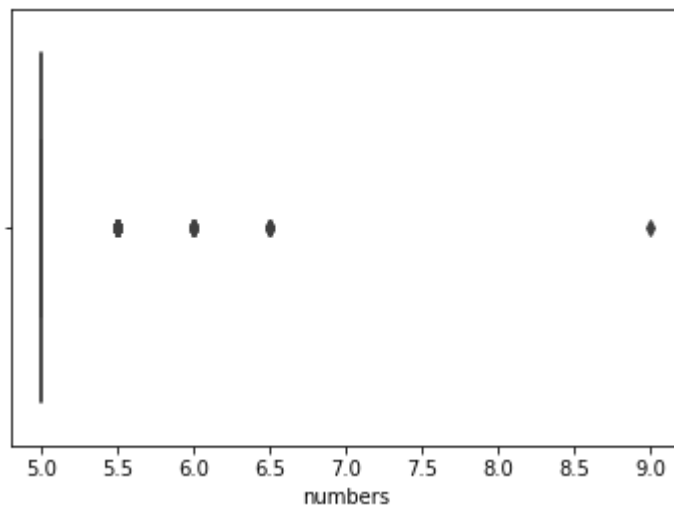
```
In [51]: sns.boxplot(x=result['PARA_A'])
```

```
Out[51]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade554940>
```



```
In [52]: sns.boxplot(x=result['numbers'])
```

```
Out[52]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade5aef28>
```



```
In [53]: result['numbers'].describe()
```

```
Out[53]: count      758.000000  
mean         5.068602  
std          0.266918  
min          5.000000  
25%          5.000000  
50%          5.000000  
75%          5.000000  
max          9.000000  
Name: numbers, dtype: float64
```

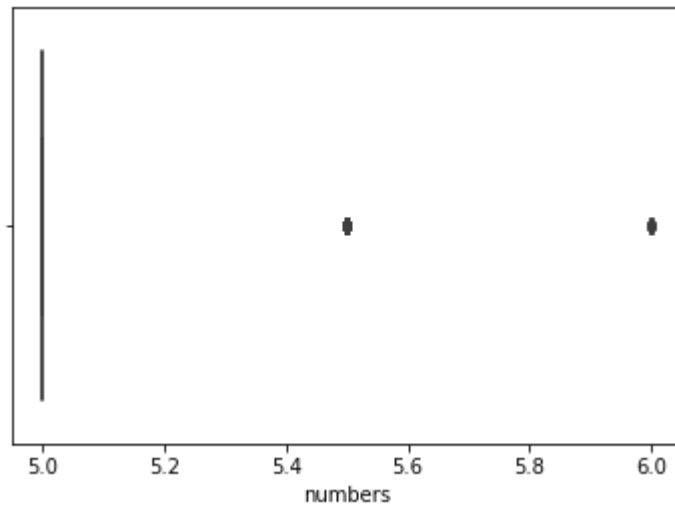
```
In [54]: i = result.loc[result['numbers'] > 6.000000].index
```

```
In [55]: result = result.drop(i)
```



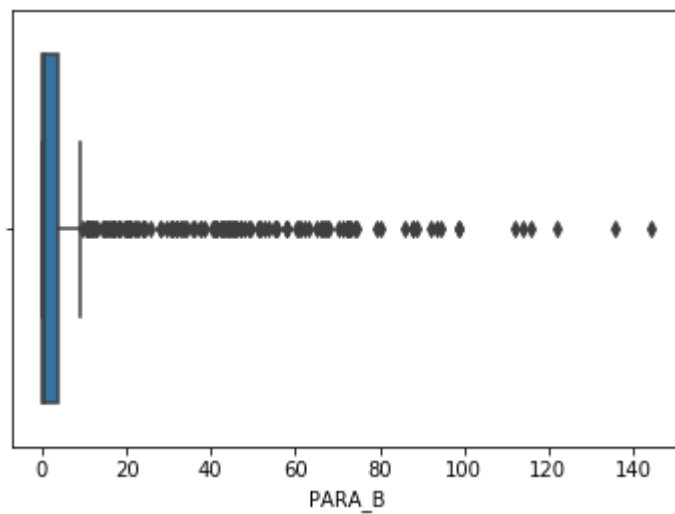
```
In [56]: sns.boxplot(x=result['numbers'])
```

```
Out[56]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade661eb8>
```



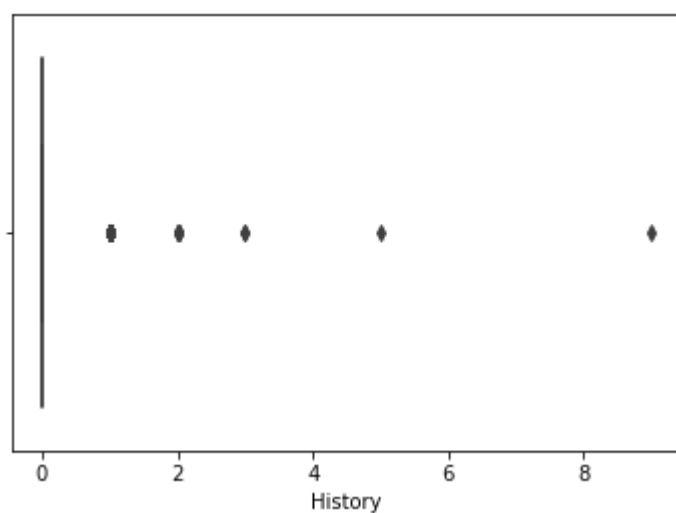
```
In [57]: sns.boxplot(x=result['PARA_B'])
```

```
Out[57]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade6ba668>
```



```
In [58]: sns.boxplot(x=result['History'])
```

```
Out[58]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade726c88>
```



```
In [59]: result['History'].describe()
```

```
Out[59]: count      750.000000
mean         0.094667
std          0.496017
min          0.000000
25%          0.000000
50%          0.000000
75%          0.000000
max          9.000000
Name: History, dtype: float64
```

```
In [60]: result.loc[result['History'] > 4.000000]
```

```
Out[60]:
```

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TOTAL
80	3.89	2	0.00	0.2	0.000	3.72	0.2	0.744	3.7
276	2.37	6	2.51	0.6	1.506	8.59	0.6	5.154	11.1

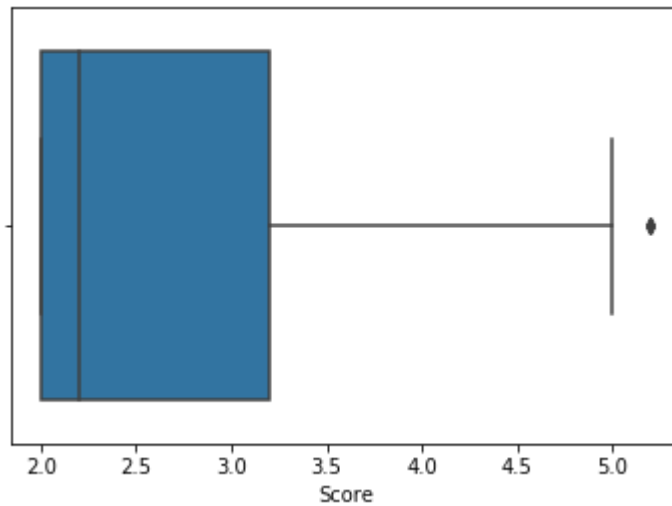
2 rows × 33 columns

```
In [61]: i = result.loc[result['History'] > 4.000000].index
```

```
In [62]: result = result.drop(i)
```

```
In [63]: sns.boxplot(x=result['Score'])
```

```
Out[63]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade77e0b8>
```



```
In [64]: result['Score'].describe()
```

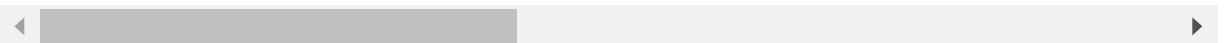
```
Out[64]: count      748.000000
mean         2.682086
std          0.834048
min          2.000000
25%          2.000000
50%          2.200000
75%          3.200000
max          5.200000
Name: Score, dtype: float64
```

```
In [65]: result.loc[result['Score'] > 5.000000]
```

```
Out[65]:
```

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TOTAL
93	3.89	16	15.72	0.6	9.432	57.94	0.6	34.764	73.61
190	2.72	18	5.11	0.6	3.066	53.76	0.6	32.256	58.8
495	15.56	1	3.73	0.6	2.238	2.82	0.6	1.692	6.5

3 rows × 33 columns

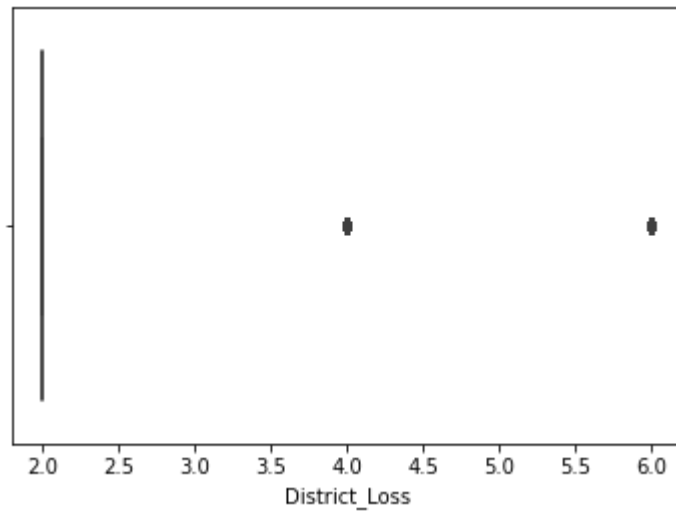


```
In [66]: i = result.loc[result['Score'] > 5.000000].index
```

```
In [67]: result = result.drop(i)
```

```
In [68]: sns.boxplot(x=result['District_Loss'])
```

```
Out[68]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade7c9cc0>
```



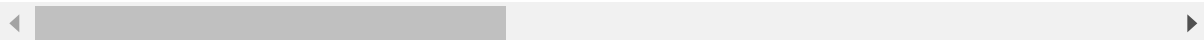
```
In [69]: result.loc[result['District_Loss'] >= 6.000000]
```

Out[69]:

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TC
94	3.89	16	4.9500	0.6	2.97000	42.61	0.6	25.566	47.
95	3.89	16	0.0000	0.2	0.00000	0.08	0.2	0.016	0.
96	3.89	16	1.1900	0.4	0.47600	12.88	0.6	7.728	14.
97	3.89	16	0.0000	0.2	0.00000	2.72	0.2	0.544	2.
98	3.89	16	0.0000	0.2	0.00000	1.66	0.2	0.332	1.
99	3.89	16	6.4700	0.6	3.88200	12.18	0.6	7.308	18.
100	3.89	16	1.3300	0.4	0.53200	0.00	0.2	0.000	1.
101	3.89	16	0.0000	0.2	0.00000	1.10	0.2	0.220	1.
125	3.41	12	25.4000	0.6	15.24000	51.74	0.6	31.044	77.
126	3.41	12	3.2100	0.6	1.92600	72.07	0.6	43.242	75.
172	3.41	12	1.0700	0.4	0.42800	0.00	0.2	0.000	1.
205	2.72	2	3.2600	0.6	1.95600	5.55	0.4	2.220	8.
242	2.72	1	4.3600	0.6	2.61600	33.91	0.6	20.346	38.
243	2.72	2	2.2300	0.6	1.33800	4.54	0.2	0.908	6.
274	2.37	1	3.7500	0.6	2.25000	1.25	0.4	0.500	5.
297	2.37	28	0.4400	0.2	0.08800	0.02	0.2	0.004	0.
298	2.37	28	1.3100	0.4	0.52400	0.12	0.2	0.024	1.
313	2.37	28	0.3200	0.2	0.06400	0.49	0.2	0.098	0.
314	2.37	28	0.5100	0.2	0.10200	0.00	0.2	0.000	0.
319	2.37	2	0.7300	0.2	0.14600	0.05	0.2	0.010	0.
320	2.37	2	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
324	2.37	2	2.1800	0.6	1.30800	0.93	0.2	0.186	3.
325	2.37	2	7.5900	0.6	4.55400	1.47	0.4	0.588	9.
326	2.37	2	0.7600	0.2	0.15200	0.34	0.2	0.068	1.
338	2.37	2	1.6600	0.4	0.66400	0.83	0.2	0.166	2.
339	2.37	2	0.5000	0.2	0.10000	0.04	0.2	0.008	0.
340	2.37	2	0.9000	0.2	0.18000	0.29	0.2	0.058	1.
341	2.37	2	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
360	1.99	20	0.8600	0.2	0.17200	0.00	0.2	0.000	0.
361	1.99	19	1.6200	0.4	0.64800	0.00	0.2	0.000	1.
...
413	1.85	19	0.1800	0.2	0.03600	0.00	0.2	0.000	0.
414	1.85	19	5.9000	0.6	3.54000	0.00	0.2	0.000	5.
415	1.85	19	4.0400	0.6	2.42400	0.00	0.2	0.000	4.
416	1.85	19	2.8100	0.6	1.68600	0.11	0.2	0.022	2.

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TC
417	1.85	19	5.8700	0.6	3.52200	0.00	0.2	0.000	5.
418	1.85	19	1.7500	0.4	0.70000	0.00	0.2	0.000	1.
419	1.85	19	2.0500	0.6	1.23000	0.00	0.2	0.000	2.
420	1.85	19	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
452	1.85	19	1.1600	0.4	0.46400	0.00	0.2	0.000	1.
453	1.85	19	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
456	1.85	19	3.2800	0.6	1.96800	0.25	0.2	0.050	3.
457	1.85	19	0.9000	0.2	0.18000	0.00	0.2	0.000	0.
477	1.85	19	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
478	1.85	19	0.9900	0.2	0.19800	0.11	0.2	0.022	1.
487	2.37	4	0.5600	0.2	0.11200	0.70	0.2	0.140	1.
488	2.37	8	0.5600	0.2	0.11200	0.08	0.2	0.016	0.
489	2.34	2	3.9500	0.6	2.37000	0.00	0.2	0.000	0.
494	15.56	2	20.6500	0.6	12.39000	0.00	0.2	0.000	0.
496	15.56	19	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
497	17.68	1	2.5200	0.6	1.51200	20.28	0.6	12.168	22.
539	59.85	1	1.9400	0.4	0.77600	6.69	0.6	4.014	8.
582	55.57	2	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
583	55.57	2	2.5900	0.6	1.55400	111.90	0.6	67.140	114.
584	55.57	2	6.4800	0.6	3.88800	43.08	0.6	25.848	49.
603	55.57	2	1.2400	0.4	0.49600	2.27	0.6	1.362	3.
604	55.57	1	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
605	55.57	44	0.0006	0.2	0.00012	1.11	0.4	0.444	1.
638	55.57	2	0.7400	0.2	0.14800	0.00	0.2	0.000	0.
639	55.57	2	0.0000	0.2	0.00000	0.00	0.2	0.000	0.
717	55.57	2	1.0600	0.4	0.42400	0.63	0.2	0.126	1.

64 rows × 33 columns



In [70]: `result['District_Loss'].describe()`

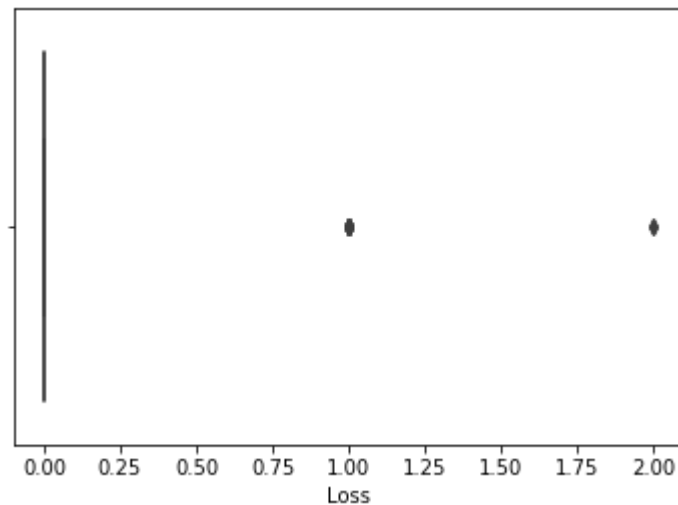
```
Out[70]: count    745.000000
mean      2.477852
std       1.190173
min       2.000000
25%      2.000000
50%      2.000000
75%      2.000000
max       6.000000
Name: District_Loss, dtype: float64
```

```
In [71]: i = result.loc[result['District_Loss'] >= 6.000000].index
```

```
In [72]: result = result.drop(i)
```

```
In [73]: sns.boxplot(x=result['Loss'])
```

```
Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0x21ade840da0>
```

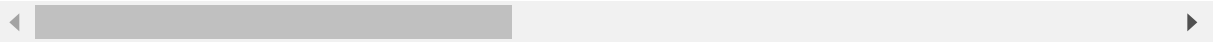



```
In [74]: result.loc[result['Loss'] >= 1.000000]
```

```
Out[74]:
```

	Sector_score	LOCATION_ID	PARA_A	Score_A	Risk_A	PARA_B	Score_B	Risk_B	TOTAL
11	3.89	8	15.38	0.6	9.228	40.14	0.6	24.084	55.5:
50	3.89	22	1.97	0.4	0.788	2.10	0.2	0.420	4.0
85	3.89	9	8.46	0.6	5.076	72.76	0.6	43.656	81.2:
121	3.41	27	1.41	0.4	0.564	36.16	0.6	21.696	37.5
143	3.41	7	6.61	0.6	3.966	55.62	0.6	33.372	62.2:
151	3.41	29	12.29	0.6	7.374	25.74	0.6	15.444	38.0:
164	3.41	19	0.31	0.2	0.062	58.12	0.6	34.872	58.4:
170	3.41	2	1.18	0.4	0.472	0.00	0.2	0.000	1.1:
206	2.72	13	3.84	0.6	2.304	144.39	0.6	86.634	148.2:
219	2.72	22	3.22	0.6	1.932	55.64	0.6	33.384	58.8:
272	2.37	18	3.43	0.6	2.058	1.00	0.4	0.400	4.4:
283	2.37	27	3.61	0.6	2.166	2.13	0.6	1.278	5.7:
391	1.85	6	8.53	0.6	5.118	0.16	0.2	0.032	8.6:
398	1.85	8	0.95	0.2	0.190	0.00	0.2	0.000	0.9:
464	1.85	22	7.84	0.6	4.704	2.81	0.6	1.686	10.6:
550	59.85	2	0.00	0.2	0.000	0.00	0.2	0.000	0.0:
647	55.57	4	0.32	0.2	0.064	0.00	0.2	0.000	0.3:

17 rows × 33 columns



```
In [75]: i = result.loc[result['Loss'] >= 1.000000].index
```

```
In [76]: result = result.drop(i)
```

```
In [77]: result['Loss'].describe()
```

```
Out[77]: count    664.0
mean         0.0
std          0.0
min          0.0
25%          0.0
50%          0.0
75%          0.0
max          0.0
Name: Loss, dtype: float64
```

```
In [78]: from scipy import stats
import numpy as np
z = np.abs(stats.zscore(result))
print(z)
```

```
[[0.70278791 0.76223977 0.73249566 ... nan 0.21419513 0.84865025]
 [0.70278791 0.94798931 0.65373269 ... nan 0.21419513 1.17834172]
 [0.70278791 0.94798931 0.48459956 ... nan 0.21419513 1.17834172]
 ...
 [1.39688538 0.14317563 0.57414063 ... nan 0.21419513 1.17834172]
 [1.39688538 0.25923122 0.58740597 ... nan 0.21419513 1.17834172]
 [1.39688538 0.04257392 0.65373269 ... nan 0.21419513 1.17834172]]
```

C:\Users\prith\Anaconda3\lib\site-packages\scipy\stats\stats.py:2253: Runtime Warning: invalid value encountered in true_divide
return (a - mns) / sstd

Regression models

Linear Regression

```
In [79]: x = result[['Sector_score', 'PARA_A', 'PARA_B', 'numbers', 'Money_Value', 'History',
'Score', 'District_Loss', 'Loss']]
y = result[['Audit_Risk']]
```

```
In [80]: from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

X_train_org, X_test_org, y_train, y_test = train_test_split(x,y, random_state
= 0) #random state

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train_org)
X_test = scaler.transform(X_test_org)
```

```
In [81]: from sklearn.linear_model import LinearRegression

lreg = LinearRegression()
lreg.fit(X_train, y_train)
print(lreg.score(X_train, y_train))
print(lreg.score(X_test, y_test))
```

```
0.941816678800454
0.9186754620340319
```

CROSS VALIDATION TECHNIQUES

```
In [82]: from sklearn.model_selection import cross_val_score

scores_tr = cross_val_score(lreg, X_train, y_train, cv=5)
scores_ts = cross_val_score(lreg, X_test, y_test, cv=5)
print("Cross-validation scores for train: {}".format(scores_tr))
print("Cross-validation scores for test : {}".format(scores_ts))
print("Average cross-validation score for train: {:.2f}".format(scores_tr.mean()))
print("Average cross-validation score for test: {:.2f}".format(scores_ts.mean()))
```

Cross-validation scores for train: [0.95024918 0.96992359 0.86595338 0.90676058 0.89215733]

Cross-validation scores for test : [0.92205629 0.85338887 0.63206558 0.86230559 0.96505314]

Average cross-validation score for train: 0.92

Average cross-validation score for test: 0.85

```
In [83]: from sklearn.model_selection import KFold

kfold = KFold(n_splits=3)

print("Cross-validation scores for train:\n{}".format(cross_val_score(lreg, X_train, y_train, cv=kfold)))
print("Cross-validation scores for test:\n{}".format(cross_val_score(lreg, X_test, y_test, cv=kfold)))
```

Cross-validation scores for train:

[0.80524208 0.82246261 0.87100622]

Cross-validation scores for test:

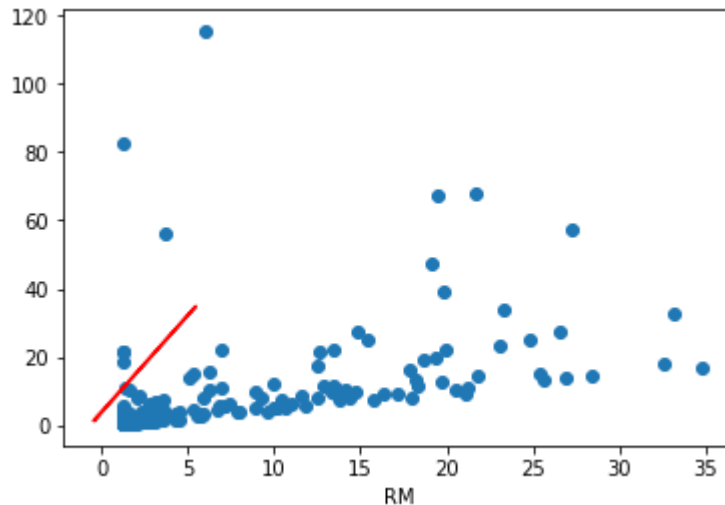
[0.84192461 0.88078059 0.86087068]

```
In [84]: %matplotlib inline
import matplotlib.pyplot as plt

X_train_rm = X_train[:,2].reshape(-1,1)
lreg.fit(X_train_rm, y_train)
y_predict = lreg.predict(X_train_rm)

plt.plot(X_train_rm, y_predict, c = 'r')
plt.scatter(y_predict,y_train)
plt.xlabel('RM')
```

Out[84]: Text(0.5,0,'RM')

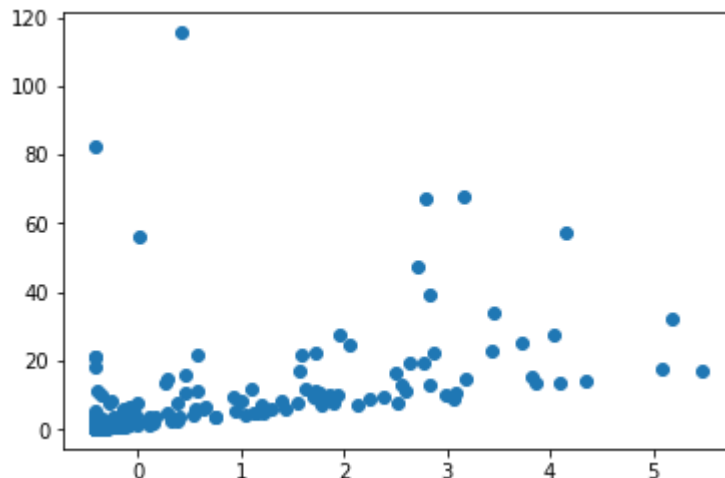


Polynomial Regression

```
In [85]: from sklearn.preprocessing import PolynomialFeatures

X_train_1 = X_train[:,2].reshape(-1,1)
plt.scatter(X_train_1,y_train)
```

Out[85]: <matplotlib.collections.PathCollection at 0x21ae094ba20>



```
In [86]: train_score_list = []
        test_score_list = []

        for n in range(1,3):
            poly = PolynomialFeatures(n)
            X_train_poly = poly.fit_transform(X_train)
            X_test_poly = poly.transform(X_test)
            lreg.fit(X_train_poly, y_train)
            train_score_list.append(lreg.score(X_train_poly, y_train))
            test_score_list.append(lreg.score(X_test_poly, y_test))
```

```
In [87]: print(train_score_list)
        print(test_score_list)

[0.941816678800454, 0.9989678528129494]
[0.9186754620340322, 0.99860111895787]
```

Cross Validation

```
In [88]: from sklearn.model_selection import cross_val_score

        scores = cross_val_score(lreg, X_train_poly, y_train, cv=5)
        print("Cross-validation scores: {}".format(scores))

Cross-validation scores: [0.93583331 0.95252906 0.87266821 0.9879375  0.99752
46 ]
```

```
In [89]: from sklearn.model_selection import KFold
        kfold = KFold(n_splits=3)

        print("Cross-validation scores for train:\n{}".format(cross_val_score(lreg, X_
train_poly, y_train, cv=kfold)))
        print("Cross-validation scores for test:\n{}".format(cross_val_score(lreg, X_t
est_poly, y_test, cv=kfold)))

Cross-validation scores for train:
[-0.0091115  0.98445248 0.99192111]
Cross-validation scores for test:
[ 0.84700854 -0.88286366 0.51469873]
```

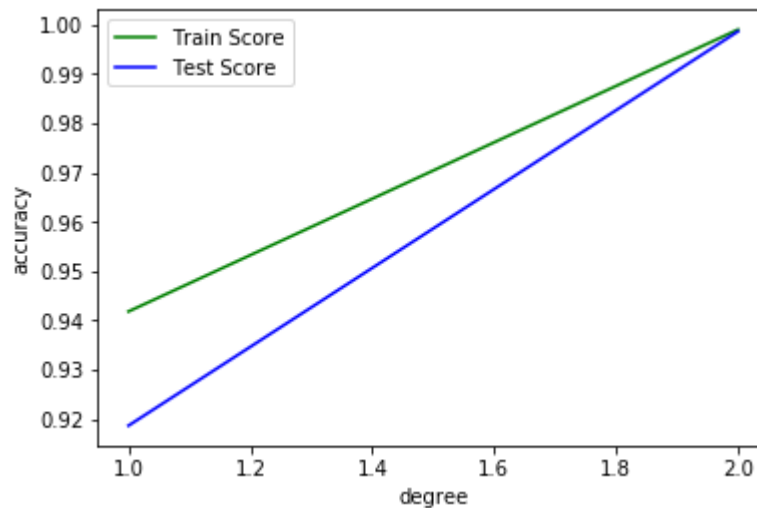
```
In [90]: print("Average cross-validation score for train: {:.2f}".format(scores.mean
()))
        print("Average cross-validation score for test: {:.2f}".format(scores_ts.mean
()))

Average cross-validation score for train: 0.95
Average cross-validation score for test: 0.85
```

```
In [91]: %matplotlib inline

x_axis = range(1,3)
plt.plot(x_axis, train_score_list, c = 'g', label = 'Train Score')
plt.plot(x_axis, test_score_list, c = 'b', label = 'Test Score')
plt.xlabel('degree')
plt.ylabel('accuracy')
plt.legend()
```

Out[91]: <matplotlib.legend.Legend at 0x21ae09b1c88>

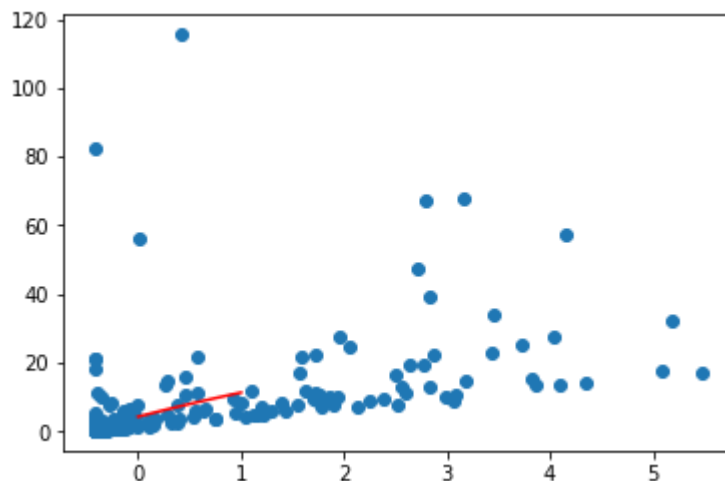


```
In [92]: poly = PolynomialFeatures(n)
X_train_poly = poly.fit_transform(X_train_1)
lreg.fit(X_train_poly, y_train)

x_axis = np.linspace(0,1,100).reshape(-1,1)
x_poly = poly.transform(x_axis)
y_predict = lreg.predict(x_poly)

X_train_1 = X_train[:,2].reshape(-1,1)
plt.scatter(X_train_1,y_train)
plt.plot(x_axis, y_predict, c = 'r')
```

Out[92]: [<matplotlib.lines.Line2D at 0x21ae1e6a828>]



SGD Regression

In []:

```
In [93]: from sklearn.linear_model import SGDRegressor

sgd_reg = SGDRegressor(random_state= 0, max_iter = 100000, learning_rate = 'optimal', penalty = 'l2')#penalty-l1-l2 #learning
sgd_reg.fit(X_train, y_train)
print(sgd_reg.score(X_train, y_train))
print(sgd_reg.score(X_test, y_test))
```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

0.9418153803190792

0.9184673967665921

GRID SEARCH

```
In [94]: from sklearn.model_selection import GridSearchCV
param_grid = {'penalty': ['l1', 'l2']}
grid_search = GridSearchCV(SGDRegressor(), param_grid, return_train_score = True)
grid_search.fit(X_train, y_train)
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:128: FutureWarning: max_iter and tol parameters have been added in
<class 'sklearn.linear_model.stochastic_gradient.SGDRegressor'> in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=1000. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.
```

```
"and default tol will be 1e-3." % type(self), FutureWarning)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:128: FutureWarning: max_iter and tol parameters have been added in
<class 'sklearn.linear_model.stochastic_gradient.SGDRegressor'> in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=1000. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.
```

```
"and default tol will be 1e-3." % type(self), FutureWarning)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:128: FutureWarning: max_iter and tol parameters have been added in
<class 'sklearn.linear_model.stochastic_gradient.SGDRegressor'> in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=1000. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.
```

```
"and default tol will be 1e-3." % type(self), FutureWarning)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:128: FutureWarning: max_iter and tol parameters have been added in
<class 'sklearn.linear_model.stochastic_gradient.SGDRegressor'> in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=1000. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.
```

```
"and default tol will be 1e-3." % type(self), FutureWarning)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:128: FutureWarning: max_iter and tol parameters have been added in
<class 'sklearn.linear_model.stochastic_gradient.SGDRegressor'> in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=1000. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.
```

```
"and default tol will be 1e-3." % type(self), FutureWarning)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:128: FutureWarning: max_iter and tol parameters have been added in
```

```
<class 'sklearn.linear_model.stochastic_gradient.SGDRegressor'> in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=1000. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.
```

```
"and default tol will be 1e-3." % type(self), FutureWarning)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:128: FutureWarning: max_iter and tol parameters have been added in <class 'sklearn.linear_model.stochastic_gradient.SGDRegressor'> in 0.19. If both are left unset, they default to max_iter=5 and tol=None. If tol is not None, max_iter defaults to max_iter=1000. From 0.21, default max_iter will be 1000, and default tol will be 1e-3.
```

```
"and default tol will be 1e-3." % type(self), FutureWarning)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
```

```
y = column_or_1d(y, warn=True)
```

```
Out[94]: GridSearchCV(cv=None, error_score='raise',
                    estimator=SGDRegressor(alpha=0.0001, average=False, epsilon=0.1, eta0=0.01,
                    fit_intercept=True, l1_ratio=0.15, learning_rate='invscaling',
                    loss='squared_loss', max_iter=None, n_iter=None, penalty='l2',
                    power_t=0.25, random_state=None, shuffle=True, tol=None, verbose=0,
                    warm_start=False),
                    fit_params=None, iid=True, n_jobs=1,
                    param_grid={'penalty': ['l1', 'l2']}, pre_dispatch='2*n_jobs',
                    refit=True, return_train_score=True, scoring=None, verbose=0)
```

```
In [95]: print("Best parameters: {}".format(grid_search.best_params_))
```

```
Best parameters: {'penalty': 'l2'}
```

CROSS VALIDATION

In [96]: `from sklearn.model_selection import cross_val_score`

```
scores = cross_val_score(sgd_reg, X_train, y_train)
print("Cross-validation scores: {}".format(scores))
```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

y = column_or_1d(y, warn=True)

Cross-validation scores: [0.80525044 0.82284828 0.87033534]

In [97]: `scores_tr = cross_val_score(sgd_reg, X_train, y_train, cv=5)`

```
print("Cross-validation scores: {}".format(scores_tr))
```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

y = column_or_1d(y, warn=True)

Cross-validation scores: [0.9506062 0.97010078 0.86596315 0.90677421 0.89187221]

```
In [98]: scores_ts = cross_val_score(sgd_reg, X_test, y_test, cv=5)
print("Cross-validation scores: {}".format(scores_ts))
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

Cross-validation scores: [0.92103594 0.85385415 0.62723122 0.86203367 0.96561
426]
```

```
In [99]: print("Average cross-validation score for train: {:.2f}".format(scores_tr.mean
()))
print("Average cross-validation score for test: {:.2f}".format(scores_ts.mean
()))
```

```
Average cross-validation score for train: 0.92
Average cross-validation score for test: 0.85
```

```
In [100]: from sklearn.model_selection import KFold
kfold = KFold(n_splits=3)

print("Cross-validation scores for train:\n{}".format(cross_val_score(sgd_reg,
X_train, y_train, cv=kfold)))
print("Cross-validation scores for test:\n{}".format(cross_val_score(sgd_reg,
X_test, y_test, cv=kfold)))
```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

Cross-validation scores for train:
[0.80525044 0.82284828 0.87033534]

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

Cross-validation scores for test:
[0.84194117 0.88096633 0.85961551]

In []:

Ridge Regresssion

```
In [101]: from sklearn.linear_model import Ridge

x_range = [0.01, 0.1, 1, 10, 100]
train_score_list = []
test_score_list = []

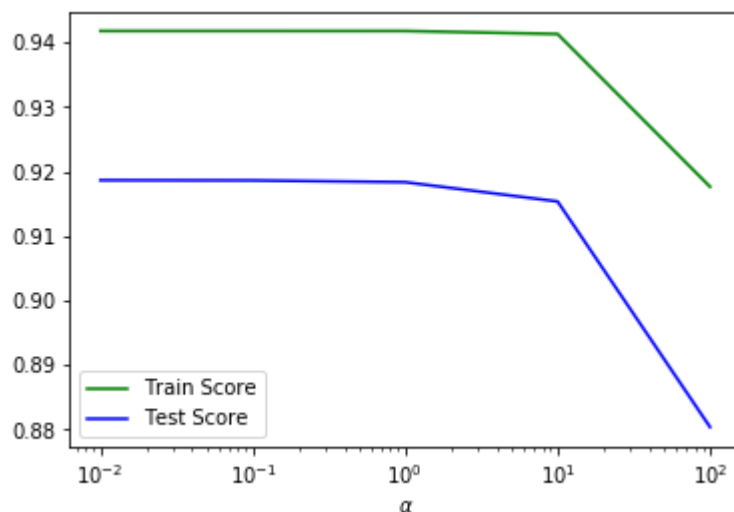
for alpha in x_range:
    ridge = Ridge(alpha)
    ridge.fit(X_train,y_train)
    train_score_list.append(ridge.score(X_train,y_train))
    test_score_list.append(ridge.score(X_test, y_test))
```

```
In [102]: print(train_score_list)
print(test_score_list)
```

```
[0.9418166782701348, 0.9418166258265133, 0.9418114386961137, 0.94134365982016
78, 0.9176719082759448]
[0.9186724376484778, 0.918645188529656, 0.9183698209065175, 0.915379354108245
3, 0.8803968403073018]
```

```
In [103]: %matplotlib inline
import matplotlib.pyplot as plt
plt.plot(x_range, train_score_list, c = 'g', label = 'Train Score')
plt.plot(x_range, test_score_list, c = 'b', label = 'Test Score')
plt.xscale('log')
plt.legend(loc = 3)
plt.xlabel(r'$\alpha$')
```

```
Out[103]: Text(0.5,0,'$\alpha$')
```



GRID SEARCH

```
In [104]: from sklearn.model_selection import GridSearchCV
param_grid = {'alpha': [0.001, 0.01, 0.1, 1, 10, 100]}
grid_search = GridSearchCV(Ridge(), param_grid, return_train_score = True)
grid_search.fit(X_train, y_train)
```

```
Out[104]: GridSearchCV(cv=None, error_score='raise',
                      estimator=Ridge(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=N
one,
                      normalize=False, random_state=None, solver='auto', tol=0.001),
                      fit_params=None, iid=True, n_jobs=1,
                      param_grid={'alpha': [0.001, 0.01, 0.1, 1, 10, 100]},
                      pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
                      scoring=None, verbose=0)
```

```
In [105]: print("Best parameters: {}".format(grid_search.best_params_))
```

```
Best parameters: {'alpha': 10}
```

```
In [106]: ridge = Ridge(alpha = 10)
ridge.fit(X_train,y_train)
print('Train score: {:.4f}'.format(ridge.score(X_train,y_train)))
print('Test score: {:.4f}'.format(ridge.score(X_test, y_test)))
```

```
Train score: 0.9413
```

```
Test score: 0.9154
```

CROSS VALIDATIONS

```
In [107]: from sklearn.model_selection import cross_val_score

scores_tr = cross_val_score(ridge, X_train, y_train, cv=5)
scores_ts = cross_val_score(ridge, X_test, y_test, cv=5)
print("Cross-validation scores for train: {}".format(scores_tr))
print("Cross-validation scores for test : {}".format(scores_ts))
print("Average cross-validation score for train: {:.2f}".format(scores_tr.mean
()))
print("Average cross-validation score for test: {:.2f}".format(scores_ts.mean
()))
```

```
Cross-validation scores for train: [0.95437112 0.97234193 0.85490566 0.905650
39 0.89666195]
```

```
Cross-validation scores for test : [0.92373856 0.85768916 0.87111574 0.736197
41 0.95447309]
```

```
Average cross-validation score for train: 0.92
```

```
Average cross-validation score for test: 0.87
```

```
In [108]: from sklearn.model_selection import KFold
kfold = KFold(n_splits=3)

print("Cross-validation scores for train:\n{}".format(cross_val_score(ridge, X_train, y_train, cv=kfold)))
print("Cross-validation scores for test:\n{}".format(cross_val_score(ridge, X_test, y_test, cv=kfold)))
```

```
Cross-validation scores for train:
[0.95284087 0.80823061 0.87778553]
Cross-validation scores for test:
[0.87335566 0.83656592 0.74007986]
```

```
In [109]: ridge.coef_
```

```
Out[109]: array([[ 0.10087059,  0.59636881,  3.36693781,  0.63435603,  6.72657366,
                   3.83647226, -0.72144601,  0.75578353,  0.          ]])
```

LASSO

```
In [110]: #on of them receive higher value, anf rest all are negligible
from sklearn.linear_model import Lasso
x_range = [0.01, 0.1, 1, 10, 100]
train_score_list = []
test_score_list = []

for alpha in x_range:
    lasso = Lasso(alpha)
    lasso.fit(X_train,y_train)
    train_score_list.append(lasso.score(X_train,y_train))
    test_score_list.append(lasso.score(X_test, y_test))
```

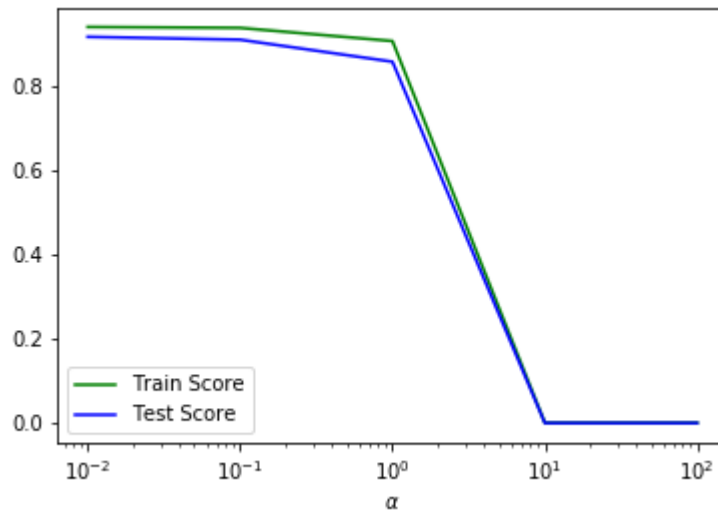
```
In [111]: print(train_score_list)
print(test_score_list)

[0.941794184561213, 0.9395986131390307, 0.9082296826545913, 0.0, 0.0]
[0.9182640925282302, 0.9114946903924728, 0.8590386572911816, -0.0007457408497821838, -0.0007457408497821838]
```



```
In [112]: #suggests smaller power of alpha is the best parameter, power of 2
plt.plot(x_range, train_score_list, c = 'g', label = 'Train Score')
plt.plot(x_range, test_score_list, c = 'b', label = 'Test Score')
plt.xscale('log')
plt.legend(loc = 3)
plt.xlabel(r'$\alpha$')
```

```
Out[112]: Text(0.5,0,'$\alpha$')
```



GRID SEARCH

```
In [113]: from sklearn.model_selection import GridSearchCV
param_grid = {'alpha': [0.001, 0.01, 0.1, 1, 10, 100]}
grid_search = GridSearchCV(Lasso(), param_grid, return_train_score = True)
grid_search.fit(X_train, y_train)

print("Best parameters: {}".format(grid_search.best_params_))
```

```
Best parameters: {'alpha': 0.1}
```

CROSS VALIDATION scores

```
In [114]: from sklearn.model_selection import cross_val_score

scores_tr = cross_val_score(lasso, X_train, y_train, cv=5)
scores_ts = cross_val_score(lasso, X_test, y_test, cv=5)
print("Cross-validation scores for train: {}".format(scores_tr))
print("Cross-validation scores for test : {}".format(scores_ts))
print("Average cross-validation score for train: {:.2f}".format(scores_tr.mean()))
print("Average cross-validation score for test: {:.2f}".format(scores_ts.mean()))
```

```
Cross-validation scores for train: [-3.79430171e-03 -2.22212602e-05 -9.539524
39e-04 -1.46277693e-02
-1.61216950e-05]
Cross-validation scores for test : [-3.12012672e-02 -2.23446537e-02 -9.708547
12e-02 -4.31044748e-03
-5.88169414e-05]
Average cross-validation score for train: -0.00
Average cross-validation score for test: -0.03
```

```
In [115]: from sklearn.model_selection import KFold

kfold = KFold(n_splits=3)

print("Cross-validation scores for train:\n{}".format(cross_val_score(lasso, X
_train, y_train, cv=kfold)))
print("Cross-validation scores for test:\n{}".format(cross_val_score(lasso, X_
test, y_test, cv=kfold)))
```

```
Cross-validation scores for train:
[-0.00450922 -0.00072537 -0.00661395]
Cross-validation scores for test:
[-2.37984342e-02 -1.56842050e-07 -5.96000207e-03]
```

VISUALIZATION

```

In [116]: #multi colinearity is not a probl in ml;
%matplotlib inline

x_range1 = np.linspace(0.001, 1, 1000).reshape(-1,1)
x_range2 = np.linspace(1, 1000, 1000).reshape(-1,1)

x_range = np.append(x_range1, x_range2)
coeff = []

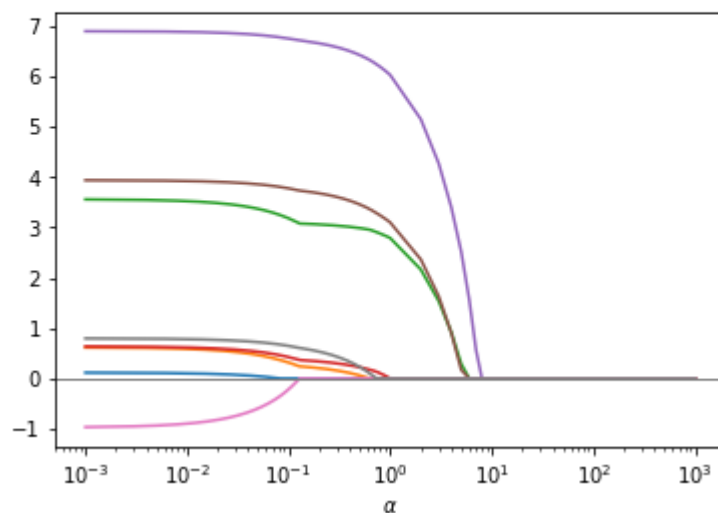
for alpha in x_range:
    lasso = Lasso(alpha)
    lasso.fit(X_train,y_train)
    coeff.append(lasso.coef_ )

coeff = np.array(coeff)

for i in range(0,8):
    plt.plot(x_range, coeff[:,i], label = 'feature {:d}'.format(i))

plt.axhline(y=0, xmin=0.001, xmax=9999, linewidth=1, c = 'gray')
plt.xlabel(r'$\alpha$')
plt.xscale('log')
plt.legend(loc='upper center', bbox_to_anchor=(0.5, 1.5),
          ncol=3, fancybox=True, shadow=True)
plt.show()


```



KNN Regressor

```
In [117]: from sklearn.neighbors import KNeighborsRegressor
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

X_train_org, X_test_org, y_train, y_test = train_test_split(x,y, random_state
= 0)

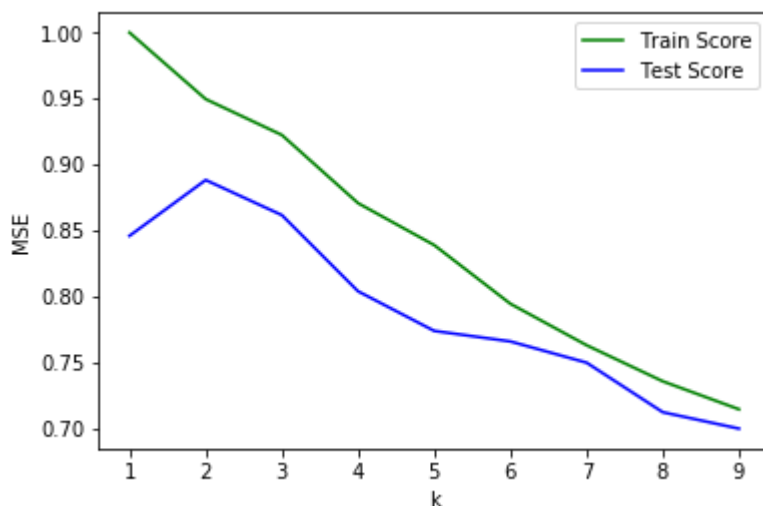
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train_org)
X_test = scaler.transform(X_test_org)
```

```
In [118]: #this generally overfits, but if still if we want, we choose 1, since atleast
#the test and train score is highest
%matplotlib inline
train_score_array = []
test_score_array = []

for k in range(1,10):
    knn_reg = KNeighborsRegressor(k)
    knn_reg.fit(X_train, y_train)
    train_score_array.append(knn_reg.score(X_train, y_train))
    test_score_array.append(knn_reg.score(X_test, y_test))

x_axis = range(1,10)
plt.plot(x_axis, train_score_array, c = 'g', label = 'Train Score')
plt.plot(x_axis, test_score_array, c = 'b', label = 'Test Score')
plt.legend()
plt.xlabel('k')
plt.ylabel('MSE')
```

Out[118]: Text(0,0.5, 'MSE')



```
In [119]: print(train_score_array)
          print(test_score_array)
```

```
[1.0, 0.949534784595555, 0.922311941343433, 0.870630995943229, 0.838804396401
6264, 0.7943045925385845, 0.7628678416982348, 0.7355136881315326, 0.714283057
6540113]
[0.845840469058989, 0.8881863460878919, 0.8615584138763992, 0.803787433942345
9, 0.7737384420856848, 0.7657481225525271, 0.7496486421464754, 0.712072699747
1046, 0.6996206815834819]
```

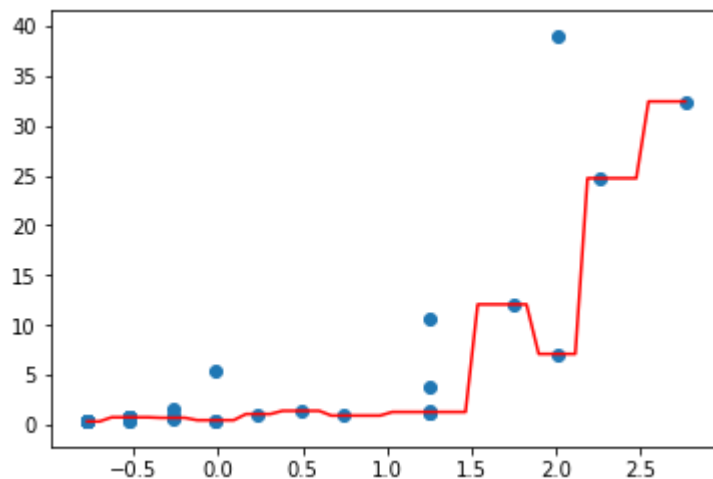
```
In [120]: X_b = X_train[:50,6].reshape(-1,1)
          y_b = y_train[:50]

          knn_reg = KNeighborsRegressor(1)
          knn_reg.fit(X_b, y_b)

          X_new=np.linspace(X_b.min(), X_b.max(), 50).reshape(50, 1)
          y_predict = knn_reg.predict(X_new)

          plt.plot(X_new, y_predict, c = 'r')
          plt.scatter(X_b, y_b)
```

```
Out[120]: <matplotlib.collections.PathCollection at 0x21ae3488278>
```



CROSS VALIDATION SCORES

```
In [121]: from sklearn.model_selection import KFold
          kfold = KFold(n_splits=5)

          print("Cross-validation scores for train:\n{}".format(cross_val_score(knn_reg,
X_train, y_train, cv=kfold)))
          print("Cross-validation scores for test:\n{}".format(cross_val_score(knn_reg,
X_test, y_test, cv=kfold)))
```

```
Cross-validation scores for train:
[0.75300686 0.76676516 0.80790769 0.86487605 0.70735587]
Cross-validation scores for test:
[0.89576307 0.83034843 0.61520757 0.3117804 0.55029848]
```

SVR

```
In [122]: from sklearn.model_selection import KFold
kfold1 = KFold(n_splits = 4, random_state=0)
from sklearn.model_selection import cross_val_score
```

```
In [123]: from sklearn.svm import SVR
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt

x = result[['Sector_score', 'PARA_A', 'PARA_B', 'numbers', 'Money_Value', 'History',
'Score', 'District_Loss', 'Loss']]
y = result[['Audit_Risk']]

X_train_org, X_test_org, y_train, y_test = train_test_split(x, y, random_state
= 0)

scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train_org)
X_test = scaler.transform(X_test_org)

svr_l = SVR(kernel = 'linear', C = 100, gamma = 10)
svr_l.fit(X_train, y_train)
print(svr_l.score(X_train, y_train))
print(svr_l.score(X_test, y_test))

0.9250577966152623
0.87754772929085
```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

```
In [124]: svr_r = SVR(kernel = 'rbf', C = 100, gamma = 10)
svr_r.fit(X_train, y_train)
print(svr_r.score(X_train, y_train))
print(svr_r.score(X_test, y_test))
```

```
0.9998495472883501
0.7689041337509936
```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

GRID SEARCH for best parameters

```
In [125]: from sklearn.svm import LinearSVR
          from sklearn.model_selection import GridSearchCV

          param_grid = {'C': [0.001, 0.01, 0.1, 1, 0.5, 10, 100], 'max_iter': [50000]}
          grid_search = GridSearchCV(LinearSVR(), param_grid, cv=kfold1, return_train_score = True)

          grid_search.fit(X_train, y_train)

          test = cross_val_score(grid_search, X_test, y_test, scoring='r2', cv=kfold1).mean()
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```



```
ed. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
```

```
In [126]: print("Train Score: {:.5f}".format(grid_search.best_score_))
print("Test Score: {:.5f}".format(test))
print("best parameters are:", grid_search.best_params_)
```

```
Train Score: 0.89523
Test Score: 0.85077
best parameters are: {'C': 100, 'max_iter': 50000}
```

```
In [127]: from sklearn.model_selection import KFold
kfold1 = KFold(n_splits = 5, random_state=0)
from sklearn.model_selection import cross_val_score
```

```
In [128]: from sklearn.svm import SVR
          from sklearn.model_selection import GridSearchCV

          param_grid = {'C': [0.001, 0.01, 0.1, 1, 0.5, 10, 100], 'max_iter': [50000]}
          grid_search = GridSearchCV(SVR(kernel='rbf'), param_grid, cv=kfold1, return_train_score = True)

          grid_search.fit(X_train, y_train)

          test = cross_val_score(grid_search, X_test, y_test, scoring='r2', cv=kfold1).mean()
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```


81/171


```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
```



```
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
```

```
In [129]: print("Train Score: {:.5f}".format(grid_search.best_score_))
          print("Test Score: {:.5f}".format(test))
          print("best parameters are:", grid_search.best_params_)
```

```
Train Score: 0.81841
Test Score: 0.77602
best parameters are: {'C': 100, 'max_iter': 50000}
```

CROSS VALIDATION

```
In [130]: from sklearn.model_selection import KFold
kfold = KFold(n_splits=3)

print("Cross-validation scores of linear for train:\n{}".format(cross_val_score(svr_l, X_train, y_train, cv=kfold)))
print("Cross-validation scores of linear for test:\n{}".format(cross_val_score(svr_l, X_test, y_test, cv=kfold)))
print("Cross-validation scores of rbf for train:\n{}".format(cross_val_score(svr_r, X_train, y_train, cv=kfold)))
print("Cross-validation scores of rbf for test:\n{}".format(cross_val_score(svr_r, X_test, y_test, cv=kfold)))
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

```
Cross-validation scores of linear for train:
[0.95926581 0.80249756 0.88150479]
Cross-validation scores of linear for test:
[0.9182423  0.80279264 0.65578351]
Cross-validation scores of rbf for train:
[0.35322915 0.47252909 0.79389827]
Cross-validation scores of rbf for test:
[0.7878341  0.34527191 0.37061812]
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Observations for Regression

Taking the case of regression, we have run each of the regression models to observe which one delivers the best score. Initially, we make use of the boxplot to check for outliers in each of the feature set values. After checking for outliers and removing them in each of the feature set variables, we find that there is not much of a case of an overfit or an underfit. The main reason for removal of outliers is to compromise on our model as a whole due to a small set of extreme values. That can be seen from each of the model scores. The best model of all the regression models is the linear regression in our case because that yields us a score of 94% in the train against a 91% in the test. None of the other models for regression came this high or close. Since, the linear regression model yields such a good score in the train and test, the other models such as the ridge, lasso and the polynomial does not matter much as they are mostly used when the linear model does not deliver a good score.

In []:

CLASSIFICATION TASK

```
In [131]: from sklearn.metrics import confusion_matrix
          from sklearn.metrics import f1_score
          from sklearn.metrics import classification_report
          from sklearn.model_selection import cross_val_score

          x = result[['Sector_score', 'PARA_A', 'PARA_B', 'numbers', 'Money_Value', 'History',
                    'Score', 'District_Loss', 'Loss']]
          y = result[['Risk']]
```

```
In [132]: X_train, X_test, y_train, y_test = train_test_split(x, y, random_state=0)
```

In []:

KNN CLASSIFIER

```
In [133]: x = result[['Sector_score', 'PARA_A', 'PARA_B', 'numbers', 'Money_Value', 'History',
                    'Score', 'District_Loss', 'Loss']]
          y = result[['Risk']]
```

```
In [134]: from sklearn.preprocessing import MinMaxScaler
          from sklearn.model_selection import train_test_split
```

```
In [135]: X_train_org, X_test_org, y_train, y_test = train_test_split(x, y, random_state
= 0, test_size = 0.2)

scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train_org)
X_test = scaler.transform(X_test_org)
```



```
In [136]: from sklearn.neighbors import KNeighborsClassifier

train_score_array = []
test_score_array = []

for k in range(1,20):
    knn = KNeighborsClassifier(k)
    knn.fit(X_train, y_train)
    train_score_array.append(knn.score(X_train, y_train))
    test_score_array.append(knn.score(X_test, y_test))
```

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

rsionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:8: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

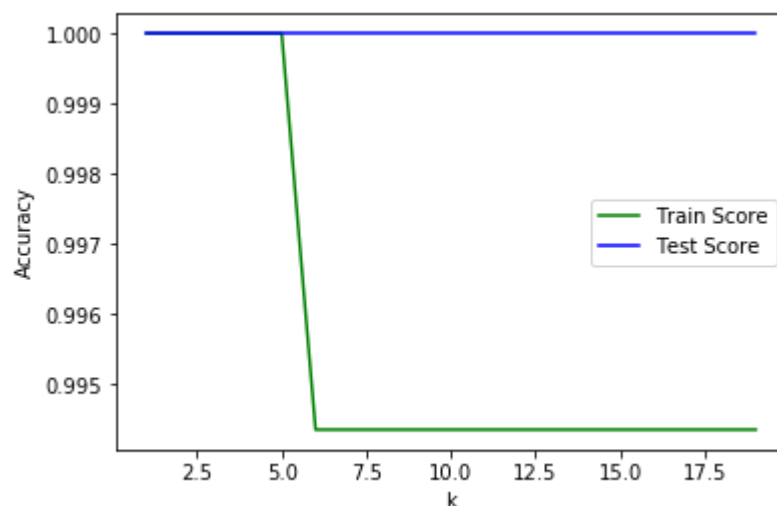
```
In [137]: print("Train score: {:.2f}".format(knn.score(X_train, y_train)))
          print("Test score: {:.2f}".format(knn.score(X_test, y_test)))
```

Train score: 0.99

Test score: 1.00

```
In [138]: #k=1 will have overfitting?/ 10 is the best: high test n train score and
          #ever similar to each other
          x_axis = range(1,20)
          %matplotlib inline
          plt.plot(x_axis, train_score_array, label = 'Train Score', c = 'g')
          plt.plot(x_axis, test_score_array, label = 'Test Score', c='b')
          plt.xlabel('k')
          plt.ylabel('Accuracy')
          plt.legend()
```

Out[138]: <matplotlib.legend.Legend at 0x21ae351ee48>



CROSS VALIDATION

```
In [139]: from sklearn.model_selection import cross_val_score

scores_tr = cross_val_score(knn, X_train, y_train, cv=5)
scores_ts = cross_val_score(knn, X_test, y_test, cv=5)
print("Cross-validation scores for train: {}".format(scores_tr))
print("Cross-validation scores for test : {}".format(scores_ts))
print("Average cross-validation score for train: {:.2f}".format(scores_tr.mean(
)))
print("Average cross-validation score for test: {:.2f}".format(scores_ts.mean(
)))
```

```

Cross-validation scores for train: [0.99065421 1.          0.9245283  1.
0.98095238]
Cross-validation scores for test : [0.92857143 0.81481481 0.92307692 0.846153
85 0.88461538]
Average cross-validation score for train: 0.98
Average cross-validation score for test: 0.88

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection\_validatio
n.py:458: DataConversionWarning: A column-vector y was passed when a 1d array
was expected. Please change the shape of y to (n_samples, ), for example usin
g ravel().

```

```

    estimator.fit(X_train, y_train, **fit_params)

```

```
In [140]: from sklearn.model_selection import KFold
kfold = KFold(n_splits=3)

print("Cross-validation scores for train:\n{}".format(cross_val_score(knn, X_train, y_train, cv=kfold)))
print("Cross-validation scores for test:\n{}".format(cross_val_score(knn, X_test, y_test, cv=kfold)))
```

Cross-validation scores for train:

[0.99435028 0.92655367 0.98870056]

Cross-validation scores for test:

[0.88888889 0.75 0.86363636]

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection_validation.py:458: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

estimator.fit(X_train, y_train, **fit_params)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection_validation.py:458: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

estimator.fit(X_train, y_train, **fit_params)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection_validation.py:458: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

estimator.fit(X_train, y_train, **fit_params)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection_validation.py:458: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

estimator.fit(X_train, y_train, **fit_params)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection_validation.py:458: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

estimator.fit(X_train, y_train, **fit_params)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\model_selection_validation.py:458: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

estimator.fit(X_train, y_train, **fit_params)

EVALUATION

```
In [141]: from sklearn.neighbors import KNeighborsClassifier
knn_e = KNeighborsClassifier().fit(X_train, y_train)
pred_knn_e = knn_e.predict(X_test)
print("Test score: {:.2f}".format(knn_e.score(X_test, y_test)))

from sklearn.metrics import confusion_matrix
print("\nKNN classifier:")
print(confusion_matrix(y_test, pred_knn_e))

from sklearn.metrics import classification_report
print(classification_report(y_test, pred_knn_e, target_names=["0", "1"]))
```

Test score: 1.00

KNN classifier:

```
[[61  0]
 [ 0 72]]
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	61
1	1.00	1.00	1.00	72
avg / total	1.00	1.00	1.00	133

C:\Users\prith\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().

SVC

```
In [142]: from sklearn.metrics import precision_recall_curve
```

```
In [143]: from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(x, y, random_state = 0)
from sklearn.svm import LinearSVC

linear_svm = LinearSVC().fit(X_train, y_train)
print("Coefficient shape: ", linear_svm.coef_.shape)
print("Intercept shape: ", linear_svm.intercept_.shape)
```

Coefficient shape: (1, 9)

Intercept shape: (1,)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)


```
In [144]: from sklearn.svm import SVC
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt

x = result[['Sector_score', 'PARA_A', 'PARA_B', 'numbers', 'Money_Value', 'History',
'Score', 'District_Loss', 'Loss']]
y = result[['Risk']]

X_train_org, X_test_org, y_train, y_test = train_test_split(x, y, random_state
= 0)

scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train_org)
X_test = scaler.transform(X_test_org)

svc_l1 = SVC(kernel = 'linear', C = 0.1, gamma = 10)
svc_l1.fit(X_train, y_train)
print(svc_l1.score(X_train, y_train))
print(svc_l1.score(X_test, y_test))

0.9317269076305221
0.9698795180722891
```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

```
In [145]: svc_r1 = SVC(kernel = 'rbf', C = 0.1, gamma = 10)
svc_r1.fit(X_train, y_train)
print(svc_r1.score(X_train, y_train))
print(svc_r1.score(X_test, y_test))

0.9317269076305221
0.9397590361445783
```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

GRID SEARCH for best paramters

```
In [146]: import numpy as np

from sklearn.svm import SVC
X_trainval, X_test, y_trainval, y_test = train_test_split(x, y, random_state=0
)
X_train, X_valid, y_train, y_valid = train_test_split(X_trainval, y_trainval,
random_state=1)

best_score = 0

for gamma in [0.001, 0.01, 0.1, 1, 10, 100]:
    for C in [0.001, 0.01, 0.1, 1, 10, 100]:
        # for each combination of parameters,
        # train an SVC
        svm = SVC(gamma=gamma, C=C)
        # perform cross-validation
        scores = cross_val_score(svm, X_trainval, y_trainval, cv=5)
        # compute mean cross-validation accuracy
        score = np.mean(scores)
        # if we got a better score, store the score and parameters
        if score > best_score:
            best_score = score
            best_parameters = {'C': C, 'gamma': gamma}
```



```
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n samples, ), for example using ravel().
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```



```
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```

In [147]: svm = SVC(**best_parameters)
          svm.fit(X_trainval, y_trainval)

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```

Out[147]: SVC(C=100, cache_size=200, class_weight=None, coef0=0.0,
              decision_function_shape='ovr', degree=3, gamma=0.1, kernel='rbf',
              max_iter=-1, probability=False, random_state=None, shrinking=True,
              tol=0.001, verbose=False)

```

```

In [148]: print("Best parameters: {}".format(grid_search.best_params_))

```

```

Best parameters: {'C': 100, 'max_iter': 50000}

```

```
In [149]: train_score = svm.score(X_train, y_train)
test_score = svm.score(X_test, y_test)
print("Train set score with best parameters: {:.2f}".format(train_score))
print("Test set score with best parameters: {:.2f}".format(test_score))
```

Train set score with best parameters: 1.00

Test set score with best parameters: 0.99


```
In [150]: from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV
param_grid = {'C': [0.001, 0.01, 0.1, 1, 10, 100],
              'gamma': [0.001, 0.01, 0.1, 1, 10, 100]}
grid_search = GridSearchCV(SVC(kernel = 'linear'), param_grid, cv=5, return_train_score=True)

X_train, X_test, y_train, y_test = train_test_split(x, y, random_state=0)
grid_search.fit(X_train, y_train)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```

```
ataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
```

```
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n samples, ), for example using ravel().
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```

```
ataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
y = column_or_1d(y, warn=True)
```



```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```

Out[150]: GridSearchCV(cv=5, error_score='raise',
    estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='auto', kernel='linear',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False),
    fit_params=None, iid=True, n_jobs=1,
    param_grid={'C': [0.001, 0.01, 0.1, 1, 10, 100], 'gamma': [0.001, 0.0
1, 0.1, 1, 10, 100]},
    pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
    scoring=None, verbose=0)

```

```
In [151]: print("Best parameters for linear SVC: {}".format(grid_search.best_params_))  
          print("Best cross-validation score: {:.2f}".format(grid_search.best_score_))
```

```
Best parameters for linear SVC: {'C': 1, 'gamma': 0.001}  
Best cross-validation score: 1.00
```

```
In [152]: from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV
param_grid = {'C': [0.001, 0.01, 0.1, 1, 10, 100],
              'gamma': [0.001, 0.01, 0.1, 1, 10, 100]}
grid_search = GridSearchCV(SVC(kernel = 'rbf'), param_grid, cv=5, return_train_score=True)

X_train, X_test, y_train, y_test = train_test_split(x, y, random_state=0)
grid_search.fit(X_train, y_train)
```


[illegible]


```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```



```
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
    y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n samples, ), for example using ravel().
```



```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)  
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().  
y = column_or_1d(y, warn=True)
```

146/171


```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```

Out[152]: GridSearchCV(cv=5, error_score='raise',
    estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False),
    fit_params=None, iid=True, n_jobs=1,
    param_grid={'C': [0.001, 0.01, 0.1, 1, 10, 100], 'gamma': [0.001, 0.0
1, 0.1, 1, 10, 100]},
    pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
    scoring=None, verbose=0)

```

```
In [153]: print("Best parameters for rbf SVC: {}".format(grid_search.best_params_))  
          print("Best cross-validation score: {:.2f}".format(grid_search.best_score_))
```

```
Best parameters for rbf SVC: {'C': 100, 'gamma': 0.1}  
Best cross-validation score: 1.00
```

```
In [ ]:
```

CROSS VALIDATION

```
In [154]: from sklearn.model_selection import KFold
kfold = KFold(n_splits=3)

print("Cross-validation scores of linear for train:\n{}".format(cross_val_score(svc_l1, X_train, y_train, cv=kfold)))
print("Cross-validation scores of linear for test:\n{}".format(cross_val_score(svc_l1, X_test, y_test, cv=kfold)))
print("Cross-validation scores of rbf for train:\n{}".format(cross_val_score(svc_r1, X_train, y_train, cv=kfold)))
print("Cross-validation scores of rbf for test:\n{}".format(cross_val_score(svc_r1, X_test, y_test, cv=kfold)))
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

Cross-validation scores of linear for train:

```
[0.97590361 0.98795181 0.98192771]
```

Cross-validation scores of linear for test:

```
[0.85714286 0.90909091 0.89090909]
```

Cross-validation scores of rbf for train:

```
[0.73493976 0.68072289 0.70481928]
```

Cross-validation scores of rbf for test:

```
[0.58928571 0.61818182 0.45454545]
```


MODEL EVALUATION using best parameters:

```
In [155]: svc_1 = SVC(kernel = 'linear', C = 0.1, gamma = 10).fit(X_train, y_train)
pred_svc_1 = svc_1.predict(X_test)
print("svc linear score: {:.2f}".format(svc_1.score(X_test, y_test)))

print("\nSVC-linear")
print(confusion_matrix(y_test, pred_svc_1))

from sklearn.metrics import classification_report
print(classification_report(y_test, pred_svc_1, target_names=["0", "1"]))
```

svc linear score: 0.99

SVC-linear

```
[[79  0]
 [ 2 85]]
```

	precision	recall	f1-score	support
0	0.98	1.00	0.99	79
1	1.00	0.98	0.99	87
avg / total	0.99	0.99	0.99	166

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

```
In [156]: svc_r = SVC(kernel = 'rbf', C = 0.1, gamma = 10).fit(X_train, y_train)
pred_svc_r = svc_r.predict(X_test)
print("svc rbf score: {:.2f}".format(svc_r.score(X_test, y_test)))

print("\nSVC-rbf")
print(confusion_matrix(y_test, pred_svc_r))

from sklearn.metrics import classification_report
print(classification_report(y_test, pred_svc_r, target_names=["0", "1"]))
```

svc rbf score: 0.70

SVC-rbf

[[29 50]

[0 87]]

	precision	recall	f1-score	support
0	1.00	0.37	0.54	79
1	0.64	1.00	0.78	87
avg / total	0.81	0.70	0.66	166

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

Precision - Recall - Linear (best score)

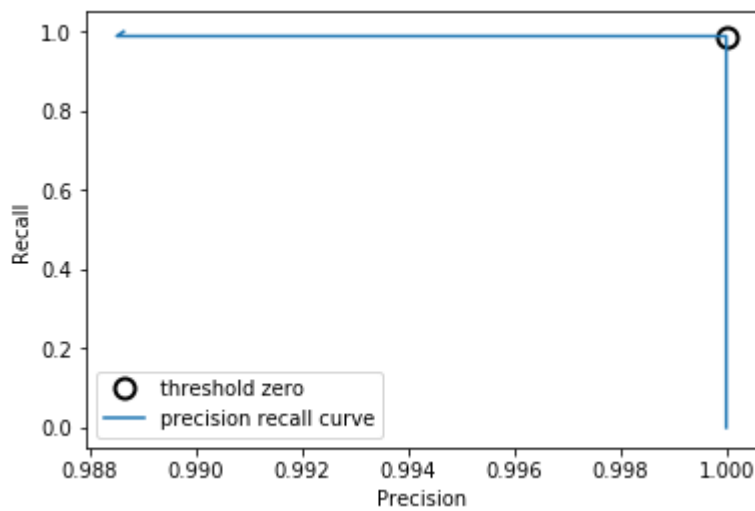
```
In [157]: %matplotlib notebook
%matplotlib inline
precision, recall, thresholds = precision_recall_curve(y_test, svc_l.decision_
function(X_test))

close_zero = np.argmin(np.abs(thresholds))

plt.plot(precision[close_zero], recall[close_zero], 'o', markersize=10,
         label="threshold zero", fillstyle="none", c='k', mew=2)

plt.plot(precision, recall, label="precision recall curve")
plt.xlabel("Precision")
plt.ylabel("Recall")
plt.legend(loc="best")
```

Out[157]: <matplotlib.legend.Legend at 0x21ae35deb00>

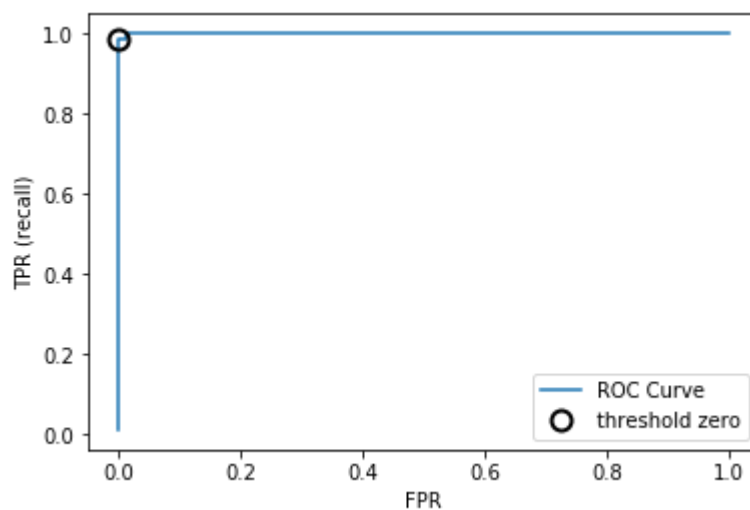


```
In [158]: %matplotlib notebook
%matplotlib inline

from sklearn.metrics import roc_curve
fpr, tpr, thresholds = roc_curve(y_test, svc_1.decision_function(X_test))

plt.plot(fpr, tpr, label="ROC Curve")
plt.xlabel("FPR")
plt.ylabel("TPR (recall)")
# find threshold closest to zero
close_zero = np.argmin(np.abs(thresholds))
plt.plot(fpr[close_zero], tpr[close_zero], 'o', markersize=10,
        label="threshold zero", fillstyle="none", c='k', mew=2)
plt.legend(loc=4)
```

Out[158]: <matplotlib.legend.Legend at 0x21ae356e898>



```
In [159]: %matplotlib notebook
          %matplotlib inline

          from sklearn.metrics import roc_auc_score

          #y = digits.target == 9

          X_train, X_test, y_train, y_test = train_test_split(
              x, y, random_state=0)

          plt.figure()

          for gamma in [1, 0.1, 0.01]:
              svc = SVC(gamma=gamma).fit(X_train, y_train)
              accuracy = svc.score(X_test, y_test)
              auc = roc_auc_score(y_test, svc.decision_function(X_test))
              fpr, tpr, _ = roc_curve(y_test, svc.decision_function(X_test))
              print("gamma = {:.2f} accuracy = {:.2f} AUC = {:.2f}".format(
                  gamma, accuracy, auc))
              plt.plot(fpr, tpr, label="gamma={:.3f}".format(gamma))
          plt.xlabel("FPR")
          plt.ylabel("TPR")
          plt.xlim(-0.01, 1)
          plt.ylim(0, 1.02)
          plt.legend(loc="best")
```

```

gamma = 1.00  accuracy = 0.99  AUC = 1.00
gamma = 0.10  accuracy = 0.99  AUC = 1.00
gamma = 0.01  accuracy = 0.96  AUC = 1.00

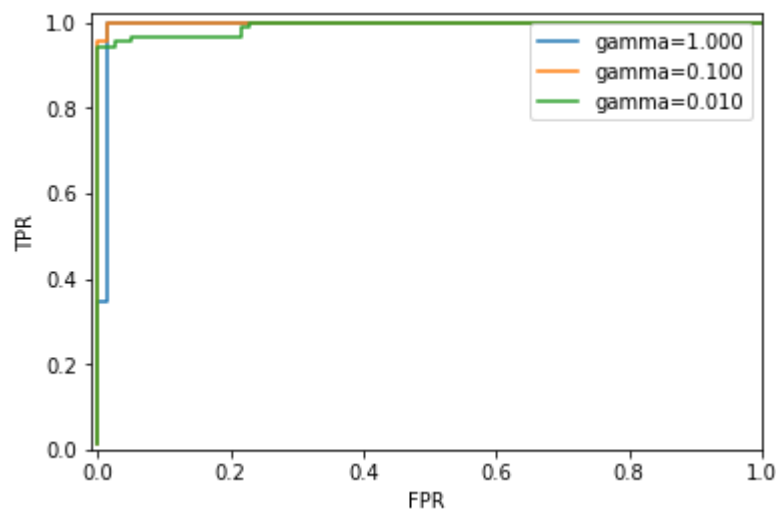
```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

Out[159]: <matplotlib.legend.Legend at 0x21ae3690828>



LOGISTIC REGRESSION

```
In [160]: import pandas as pd
```

```
In [161]: x = result[['Sector_score', 'PARA_A', 'PARA_B', 'numbers', 'Money_Value', 'History',
'Score', 'District_Loss', 'Loss']]
y = result[['Risk']]
```

```
In [162]: from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split

X_train_org, X_test_org, y_train, y_test = train_test_split(x,y, random_state
= 0)

scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train_org)
X_test = scaler.transform(X_test_org)
```

```
In [163]: from sklearn.linear_model import LogisticRegression

c_range = [0.001, 0.01, 0.1, 1, 10, 100, 1000]
train_score_l1 = []
train_score_l2 = []
test_score_l1 = []
test_score_l2 = []

for c in c_range:
    log_l1 = LogisticRegression(penalty = 'l1', C = c)
    log_l2 = LogisticRegression(penalty = 'l2', C = c)
    log_l1.fit(X_train, y_train)
    log_l2.fit(X_train, y_train)
    train_score_l1.append(log_l1.score(X_train, y_train))
    train_score_l2.append(log_l2.score(X_train, y_train))
    test_score_l1.append(log_l1.score(X_test, y_test))
    test_score_l2.append(log_l2.score(X_test, y_test))
```



```
In [164]: print(train_score_l1, train_score_l2)
          print(test_score_l1, test_score_l2)

[0.39959839357429716, 0.39959839357429716, 0.9457831325301205, 1.0, 1.0, 1.0,
1.0] [0.6004016064257028, 0.6385542168674698, 0.7751004016064257, 0.959839357
4297188, 0.9738955823293173, 1.0, 1.0]
[0.4759036144578313, 0.4759036144578313, 0.9698795180722891, 1.0, 1.0, 1.0,
1.0] [0.5240963855421686, 0.5843373493975904, 0.7951807228915663, 0.969879518
0722891, 0.9759036144578314, 1.0, 1.0]
```

GRID SEARCH for best parameters

```
In [165]: from sklearn.model_selection import GridSearchCV
param_grid = { 'penalty': ['l1', 'l2']}
grid_search = GridSearchCV(LogisticRegression(), param_grid, return_train_score = True)
grid_search.fit(X_train, y_train)
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

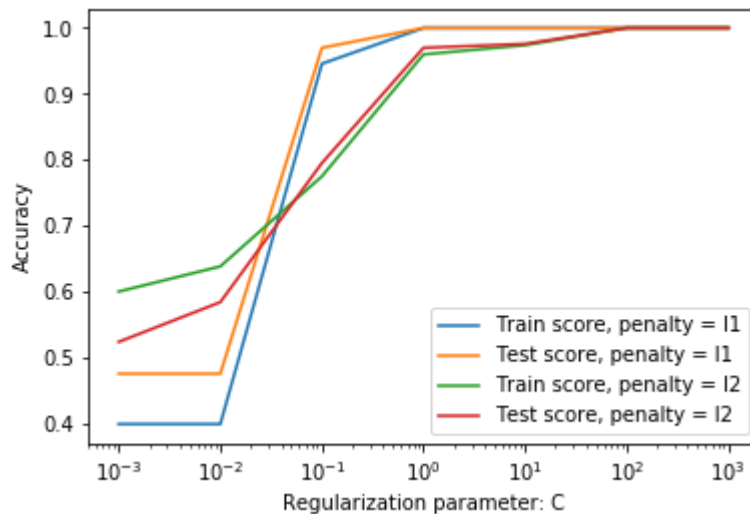
```
Out[165]: GridSearchCV(cv=None, error_score='raise',
  estimator=LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
    intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
    penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
    verbose=0, warm_start=False),
  fit_params=None, iid=True, n_jobs=1,
  param_grid={'penalty': ['l1', 'l2']}, pre_dispatch='2*n_jobs',
  refit=True, return_train_score=True, scoring=None, verbose=0)
```

```
In [166]: print("Best parameters for linear logistic Regression: {}".format(grid_search.best_params_))
```

```
Best parameters for linear logistic Regression: {'penalty': 'l1'}
```

```
In [167]: import matplotlib.pyplot as plt
%matplotlib inline

plt.plot(c_range, train_score_l1, label = 'Train score, penalty = l1')
plt.plot(c_range, test_score_l1, label = 'Test score, penalty = l1')
plt.plot(c_range, train_score_l2, label = 'Train score, penalty = l2')
plt.plot(c_range, test_score_l2, label = 'Test score, penalty = l2')
plt.legend()
plt.xlabel('Regularization parameter: C')
plt.ylabel('Accuracy')
plt.xscale('log')
```



```
In [168]: from sklearn.linear_model import LogisticRegression

log1 = LogisticRegression(penalty = 'l2', random_state = 0)
log1.fit(X_train,y_train)

print("logreg score: {:.2f}".format(log1.score(X_test, y_test)))
print("logreg train score: {:.2f}".format(log1.score(X_train, y_train)))
```

```
logreg score: 0.97
logreg train score: 0.96
```

```
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

CROSS VALIDATIONS

```
In [169]: from sklearn.model_selection import cross_val_score

scores_tr = cross_val_score(log1, X_train, y_train, cv=5)
scores_ts = cross_val_score(log1, X_test, y_test, cv=5)
print("Cross-validation scores for train: {}".format(scores_tr))
print("Cross-validation scores for test : {}".format(scores_ts))
print("Average cross-validation score for train: {:.2f}".format(scores_tr.mean(
)))
print("Average cross-validation score for test: {:.2f}".format(scores_ts.mean(
)))
```

```

Cross-validation scores for train: [0.97      0.95      0.92      0.96
0.98979592]
Cross-validation scores for test : [0.91176471 0.91176471 0.96969697 0.939393
94 0.9375   ]
Average cross-validation score for train: 0.96
Average cross-validation score for test: 0.93

```

```

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: D
ataConversionWarning: A column-vector y was passed when a 1d array was expect
ed. Please change the shape of y to (n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)

```

```
In [170]: from sklearn.model_selection import KFold
kfold = KFold(n_splits=3)

print("Cross-validation scores for train:\n{}".format(cross_val_score(log1, X_train, y_train, cv=kfold)))
print("Cross-validation scores for test:\n{}".format(cross_val_score(log1, X_test, y_test, cv=kfold)))
```

Cross-validation scores for train:

[0.96987952 0.91566265 0.97590361]

Cross-validation scores for test:

[0.875 0.87272727 0.98181818]

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

MODEL EVALUATION

```
In [171]: logreg = LogisticRegression(C=1, penalty = 'l2').fit(X_train, y_train)
pred_logreg = logreg.predict(X_test)
print("logreg score: {:.2f}".format(logreg.score(X_test, y_test)))

from sklearn.metrics import confusion_matrix
print("\nLogistic Regression")
print(confusion_matrix(y_test, pred_logreg))

from sklearn.metrics import classification_report
print(classification_report(y_test, pred_logreg, target_names=["0", "1"]))
```

logreg score: 0.97

Logistic Regression

[[79 0]

[5 82]]

		precision	recall	f1-score	support
	0	0.94	1.00	0.97	79
	1	1.00	0.94	0.97	87
avg / total		0.97	0.97	0.97	166

C:\Users\prith\Anaconda3\lib\site-packages\sklearn\utils\validation.py:578: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

DECISION TREE

```
In [172]: %matplotlib notebook
from sklearn.tree import DecisionTreeClassifier

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(x, y, stratify=y, random_state=0)

dtree = DecisionTreeClassifier(random_state=0)
#goes full length; no early stopping
dtree.fit(X_train, y_train)
print("Accuracy on training set: {:.3f}".format(dtree.score(X_train, y_train)))
print("Accuracy on test set: {:.3f}".format(dtree.score(X_test, y_test)))
```

Accuracy on training set: 1.000

Accuracy on test set: 1.000

Grid Search for best parameters

```
In [173]: from sklearn.model_selection import GridSearchCV
param_grid = {'max_depth': [1, 2, 3, 4, 5, 6]}
grid_search = GridSearchCV(DecisionTreeClassifier(), param_grid, return_train_score = True)
grid_search.fit(X_train, y_train)
```

```
Out[173]: GridSearchCV(cv=None, error_score='raise',
      estimator=DecisionTreeClassifier(class_weight=None, criterion='gini',
      max_depth=None,
      max_features=None, max_leaf_nodes=None,
      min_impurity_decrease=0.0, min_impurity_split=None,
      min_samples_leaf=1, min_samples_split=2,
      min_weight_fraction_leaf=0.0, presort=False, random_state=None,
      splitter='best'),
      fit_params=None, iid=True, n_jobs=1,
      param_grid={'max_depth': [1, 2, 3, 4, 5, 6]},
      pre_dispatch='2*n_jobs', refit=True, return_train_score=True,
      scoring=None, verbose=0)
```

```
In [174]: print("best parameters are:", grid_search.best_params_)

best parameters are: {'max_depth': 1}
```

```
In [175]: dtree = DecisionTreeClassifier(max_depth=4, random_state=0) #max depth is early stopping
dtree.fit(X_train, y_train)

print("Accuracy on training set: {:.3f}".format(dtree.score(X_train, y_train)))
print("Accuracy on test set: {:.3f}".format(dtree.score(X_test, y_test)))

Accuracy on training set: 1.000
Accuracy on test set: 1.000
```

CROSS VALIDATIONS

```
In [176]: from sklearn.model_selection import cross_val_score

scores_tr = cross_val_score(dtrees, X_train, y_train, cv=5)
scores_ts = cross_val_score(dtrees, X_test, y_test, cv=5)
print("Cross-validation scores for train: {}".format(scores_tr))
print("Cross-validation scores for test : {}".format(scores_ts))
print("Average cross-validation score for train: {:.2f}".format(scores_tr.mean()))
print("Average cross-validation score for test: {:.2f}".format(scores_ts.mean()))
```

```
Cross-validation scores for train: [1. 1. 1. 1. 1.]
Cross-validation scores for test : [1. 1. 1. 1. 1.]
Average cross-validation score for train: 1.00
Average cross-validation score for test: 1.00
```

```
In [177]: from sklearn.model_selection import KFold

kfold = KFold(n_splits=3)

print("Cross-validation scores for train:\n{}".format(cross_val_score(dtrees, X_train, y_train, cv=kfold)))
print("Cross-validation scores for test:\n{}".format(cross_val_score(dtrees, X_test, y_test, cv=kfold)))
```

```
Cross-validation scores for train:
[1. 1. 1.]
Cross-validation scores for test:
[1. 1. 1.]
```

MODEL EVALUATION using best parameters

```
In [178]: from sklearn.tree import DecisionTreeClassifier
tree = DecisionTreeClassifier(max_depth=1).fit(X_train, y_train)
pred_tree = tree.predict(X_test)
print("Test score: {:.2f}".format(tree.score(X_test, y_test)))

from sklearn.metrics import confusion_matrix
print("\nDecision tree:")
print(confusion_matrix(y_test, pred_tree))

from sklearn.metrics import classification_report
print(classification_report(y_test, pred_tree, target_names=["0", "1"]))
```

Test score: 1.00

Decision tree:

```
[[70  0]
 [ 0 96]]
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	70
1	1.00	1.00	1.00	96
avg / total	1.00	1.00	1.00	166

Observations for Classification

A good evaluation strategy in this case would be the precision because while comparing each of the strategies for each model, we find the precision to yield the best score.

The decision tree model yields the best result and in this case, yields a perfect score. The decision tree is considered one of the best models because of the ease of interpretation and also due to the ability to map non-linear relationships as well. Decision tree works for both categorical as well as continuous variables and due to the overall feature set selection again could be one of the main contributing factors for the scores. Best parameter for Max_depth of a decision tree comes out to be 1 using the grid search

In []: