

Report Q3

Dataset: Microsoft learning to Rank dataset

Reading the dataset using pandas dataframe.

Filtering the data with qid values = 4 only and removing all remaining rows.

For max order DCG:

Sort the dataset in descending order of column 0 i.e labels of the Urls.

Save the dataset in file “maxDCG.txt”.

No of such files:

19893497375938370599826047614905329896936840170566570588205180312704857992
6951934824126865654310502400000000000000000000

```
ndcg(50) = 10.323516383590077
```

```
ndcg (entire dataset) = 28.98846753873482
```

Ratio = 0.35612494416255847

For urls based on feature_75

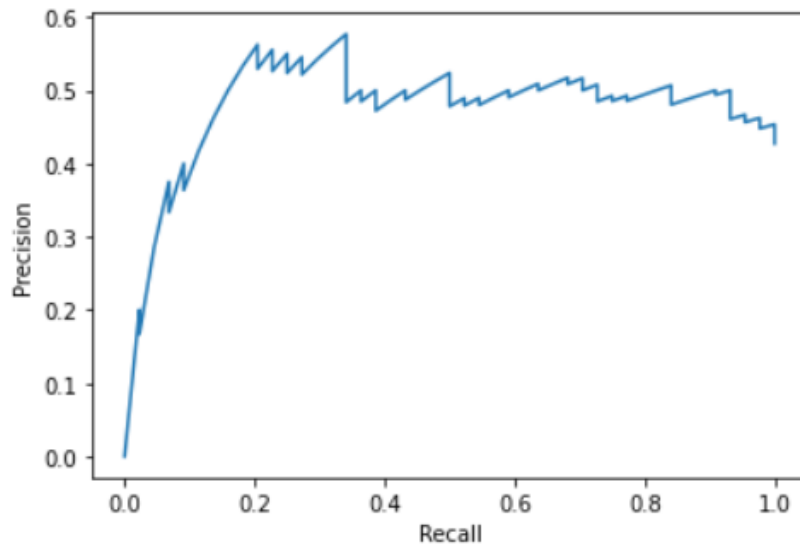
Sort the labels based on decreasing order of feature_75 using two lists :

```
arr_1 =[list of labels]
```

```
arr_2=[list of feature_75]
```

Sort using zipping the two list.

Precision-Recall Curve



From the graph , it is clear that initially the precision increases quickly for lower value of recall and then reaches threshold value after with saturation is achieved with slight variation continuing.