DataAnalysis

June 29, 2020

[69]:
```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.model_selection import train_test_split
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[70]:
```
PATH =   r"/home/swastik/workspace/repo/PyResearch/training/"
PATHTest =   r"/home/swastik/workspace/repo/PyResearch/testing/"

rawdata = pd.read_csv(PATHTest+"part-067.csv")
splitData = train_test_split(rawdata, test_size= 0.3)

train = splitData[0]
test = splitData[1]

np.shape(test)
```

[70]:
```
(12076, 235)
```

[71]:
```
df.isAnomaly.value_counts()
```

[71]:
```
False    40231
True        21
Name: isAnomaly, dtype: int64
```

[72]:
```
df[df['isAnomaly']==True].head()
```

[72]:
```
          host process         timestamp  isAnomaly  \
9795  lphost09    wls1  2015-02-09 17:00       True
9796  lphost09    wls1  2015-02-09 17:01       True
9797  lphost09    wls1  2015-02-09 17:02       True
9798  lphost09    wls1  2015-02-09 17:03       True
9799  lphost09    wls1  2015-02-09 17:04       True

      Prepared statement cache hit rate :
((MXBean(com.bea:Name=source06,Type=JDBCDataSourceRuntime).PrepStmtCacheHitCount
/ MXBean(com.bea:Name=source06,Type=JDBCDataSourceRuntime).PrepStmtCacheMissCoun
t))  \
9795                                                3.0
9796                                                1.0
9797                                                1.0
9798                                                1.0
9799                                                1.0

      Memory space usage : ((MXBean(java.lang:name=Code
Cache,type=MemoryPool).Usage.committed / MXBean(java.lang:name=Code
Cache,type=MemoryPool).Usage.max))  \
9795                                           0.981771
9796                                           0.981771
9797                                           0.981771
9798                                           0.981771
9799                                           0.981771

      Active connections : (MXBean(com.bea:Name=source04,Type=JDBCConnectionPool
Runtime).ActiveConnectionsCurrentCount)  \
9795                                                0.0
9796                                                0.0
9797                                                0.0
9798                                                0.0
9799                                                0.0

      Available db connection activity : (d/dx
(MXBean(com.bea:Name=source02,Type=JDBCDataSourceRuntime).NumAvailable))  \
9795                                                0.0
9796                                                0.0
9797                                                0.0
9798                                                0.0
9799                                                0.0

      Active connections : (MXBean(com.bea:Name=source03,Type=JDBCConnectionPool
Runtime).ActiveConnectionsCurrentCount)  \
9795                                                0.0
9796                                                0.0
9797                                                0.0
9798                                                0.0
9799                                                0.0

      DB connection started : (incld/dx (MXBean(com.bea:Name=source02,Type=JDBCD
ataSourceRuntime).ConnectionsTotalCount))  \
9795                                                0.0
9796                                                0.0
9797                                                0.0
9798                                                0.0
```

```
9799                                                 0.0

      …  \
9795  …
9796  …
9797  …
9798  …
9799  …

      Available db connection activity : (d/dx
(MXBean(com.bea:Name=source08,Type=JDBCDataSourceRuntime).NumAvailable))  \
9795                                                 0.0
9796                                                 0.0
9797                                                 0.0
9798                                                 0.0
9799                                                 0.0

      Available db connection activity : (d/dx
(MXBean(com.bea:Name=source10,Type=JDBCConnectionPoolRuntime).NumAvailable))  \
9795                                                 0.0
9796                                                 0.0
9797                                                 0.0
9798                                                 0.0
9799                                                 0.0

      Rel. unavailable connections :
((MXBean(com.bea:Name=source04,Type=JDBCDataSourceRuntime).NumUnavailable /
MXBean(com.bea:Name=source04,Type=JDBCDataSourceRuntime).CurrCapacity))  \
9795                                                 0.0
9796                                                 0.0
9797                                                 0.0
9798                                                 0.0
9799                                                 0.0

      Failed wait for connection : (incld/dx (MXBean(com.bea:Name=source08,Type=
JDBCDataSourceRuntime).WaitingForConnectionFailureTotal))  \
9795                                                 0.0
9796                                                 0.0
9797                                                 0.0
9798                                                 0.0
9799                                                 0.0

      Rel. unavailable connections :
((MXBean(com.bea:Name=source08,Type=JDBCDataSourceRuntime).NumUnavailable /
MXBean(com.bea:Name=source08,Type=JDBCDataSourceRuntime).CurrCapacity))  \
9795                                                 1.0
9796                                                 1.0
```

```
9797                                                1.0
9798                                                1.0
9799                                                1.0

      Stuck threads : (MXBean(com.bea:ApplicationRuntime=source05,Name=default,T
ype=WorkManagerRuntime).StuckThreadCount)  \
9795                                                0.0
9796                                                0.0
9797                                                0.0
9798                                                0.0
9799                                                0.0

      Process CPU : (\Process(java)\CPU)  \
9795                                 0.0
9796                                 0.0
9797                                 0.0
9798                                 0.0
9799                                 0.0

      Successful wait for connection : (incld/dx (MXBean(com.bea:Name=source03,T
ype=JDBCDataSourceRuntime).WaitingForConnectionSuccessTotal))  \
9795                                                0.0
9796                                                0.0
9797                                                0.0
9798                                                0.0
9799                                                0.0

      Failed wait for connection : (incld/dx (MXBean(com.bea:Name=source03,Type=
JDBCDataSourceRuntime).WaitingForConnectionFailureTotal))  \
9795                                                0.0
9796                                                0.0
9797                                                0.0
9798                                                0.0
9799                                                0.0

      Connection delay :
(MXBean(com.bea:Name=source02,Type=JDBCDataSourceRuntime).ConnectionDelayTime)
9795                                               90.0
9796                                               90.0
9797                                               90.0
9798                                               90.0
9799                                               90.0

[5 rows x 235 columns]
```

[73]:

```
df[df['isAnomaly']==True].describe()
```

[73]:

```
       Prepared statement cache hit rate :
((MXBean(com.bea:Name=source06,Type=JDBCDataSourceRuntime).PrepStmtCacheHitCount
/ MXBean(com.bea:Name=source06,Type=JDBCDataSourceRuntime).PrepStmtCacheMissCoun
t))  \
count                                                 21.0
mean                                                   1.0
std                                                    0.0
min                                                    1.0
25%                                                    1.0
50%                                                    1.0
75%                                                    1.0
max                                                    1.0

       Memory space usage : ((MXBean(java.lang:name=Code
Cache,type=MemoryPool).Usage.committed / MXBean(java.lang:name=Code
Cache,type=MemoryPool).Usage.max))  \
count                                         2.100000e+01
mean                                          9.817708e-01
std                                           2.275280e-16
min                                           9.817708e-01
25%                                           9.817708e-01
50%                                           9.817708e-01
75%                                           9.817708e-01
max                                           9.817708e-01

       Active connections : (MXBean(com.bea:Name=source04,Type=JDBCConnectionPoo
lRuntime).ActiveConnectionsCurrentCount)  \
count                                                 21.0
mean                                                   0.0
std                                                    0.0
min                                                    0.0
25%                                                    0.0
50%                                                    0.0
75%                                                    0.0
max                                                    0.0

       Available db connection activity : (d/dx
(MXBean(com.bea:Name=source02,Type=JDBCDataSourceRuntime).NumAvailable))  \
count                                                 21.0
mean                                                   0.0
std                                                    0.0
min                                                    0.0
25%                                                    0.0
50%                                                    0.0
75%                                                    0.0
max                                                    0.0
```

```
       Active connections : (MXBean(com.bea:Name=source03,Type=JDBCConnectionPoo
lRuntime).ActiveConnectionsCurrentCount)  \
count                                                21.0
mean                                                  0.0
std                                                   0.0
min                                                   0.0
25%                                                   0.0
50%                                                   0.0
75%                                                   0.0
max                                                   0.0

       DB connection started : (incld/dx (MXBean(com.bea:Name=source02,Type=JDBC
DataSourceRuntime).ConnectionsTotalCount))  \
count                                                21.0
mean                                                  0.0
std                                                   0.0
min                                                   0.0
25%                                                   0.0
50%                                                   0.0
75%                                                   0.0
max                                                   0.0

       Heap usage activity : (d/dx
(MXBean(java.lang:type=Memory).HeapMemoryUsage.used))  \
count                                        2.100000e+01
mean                                         2.498082e+07
std                                          8.151607e+07
min                                         -3.991606e+07
25%                                         -6.353316e+06
50%                                          1.475819e+07
75%                                          1.606932e+07
max                                          3.698590e+08

       Connection delay :
(MXBean(com.bea:Name=source10,Type=JDBCDataSourceRuntime).ConnectionDelayTime)
\
count                                                21.0
mean                                                120.0
std                                                   0.0
min                                                 120.0
25%                                                 120.0
50%                                                 120.0
75%                                                 120.0
max                                                 120.0

       Stuck threads :
(MXBean(com.bea:Name=ThreadPoolRuntime,Type=ThreadPoolRuntime).StuckThreadCount)
```

```
\
count                                               21.0
mean                                                 0.0
std                                                  0.0
min                                                  0.0
25%                                                  0.0
50%                                                  0.0
75%                                                  0.0
max                                                  0.0

       Reserve request activity : (incld/dx
(MXBean(com.bea:Name=source02,Type=JDBCDataSourceRuntime).ReserveRequestCount))
\
count                                               21.0
mean                                                 0.0
std                                                  0.0
min                                                  0.0
25%                                                  0.0
50%                                                  0.0
75%                                                  0.0
max                                                  0.0

       …  \
count  …
mean   …
std    …
min    …
25%    …
50%    …
75%    …
max    …

       Available db connection activity : (d/dx
(MXBean(com.bea:Name=source08,Type=JDBCDataSourceRuntime).NumAvailable))  \
count                                               21.0
mean                                                 0.0
std                                                  0.0
min                                                  0.0
25%                                                  0.0
50%                                                  0.0
75%                                                  0.0
max                                                  0.0

       Available db connection activity : (d/dx
(MXBean(com.bea:Name=source10,Type=JDBCConnectionPoolRuntime).NumAvailable))  \
count                                               21.0
mean                                                 0.0
```

```
std                                                    0.0
min                                                    0.0
25%                                                    0.0
50%                                                    0.0
75%                                                    0.0
max                                                    0.0

       Rel. unavailable connections :
((MXBean(com.bea:Name=source04,Type=JDBCDataSourceRuntime).NumUnavailable /
MXBean(com.bea:Name=source04,Type=JDBCDataSourceRuntime).CurrCapacity))  \
count                                                 21.0
mean                                                   0.0
std                                                    0.0
min                                                    0.0
25%                                                    0.0
50%                                                    0.0
75%                                                    0.0
max                                                    0.0

       Failed wait for connection : (incld/dx (MXBean(com.bea:Name=source08,Type
=JDBCDataSourceRuntime).WaitingForConnectionFailureTotal))  \
count                                                 21.0
mean                                                   0.0
std                                                    0.0
min                                                    0.0
25%                                                    0.0
50%                                                    0.0
75%                                                    0.0
max                                                    0.0

       Rel. unavailable connections :
((MXBean(com.bea:Name=source08,Type=JDBCDataSourceRuntime).NumUnavailable /
MXBean(com.bea:Name=source08,Type=JDBCDataSourceRuntime).CurrCapacity))  \
count                                                 21.0
mean                                                   1.0
std                                                    0.0
min                                                    1.0
25%                                                    1.0
50%                                                    1.0
75%                                                    1.0
max                                                    1.0

       Stuck threads : (MXBean(com.bea:ApplicationRuntime=source05,Name=default,
Type=WorkManagerRuntime).StuckThreadCount)  \
count                                                 21.0
mean                                                   0.0
std                                                    0.0
```

```
min                                                  0.0
25%                                                  0.0
50%                                                  0.0
75%                                                  0.0
max                                                  0.0

       Process CPU : (\Process(java)\CPU)  \
count                           21.000000
mean                             0.380952
std                              0.804748
min                              0.000000
25%                              0.000000
50%                              0.000000
75%                              0.000000
max                              2.000000

       Successful wait for connection : (incld/dx (MXBean(com.bea:Name=source03,
Type=JDBCDataSourceRuntime).WaitingForConnectionSuccessTotal))  \
count                                               21.0
mean                                                 0.0
std                                                  0.0
min                                                  0.0
25%                                                  0.0
50%                                                  0.0
75%                                                  0.0
max                                                  0.0

       Failed wait for connection : (incld/dx (MXBean(com.bea:Name=source03,Type
=JDBCDataSourceRuntime).WaitingForConnectionFailureTotal))  \
count                                               21.0
mean                                                 0.0
std                                                  0.0
min                                                  0.0
25%                                                  0.0
50%                                                  0.0
75%                                                  0.0
max                                                  0.0

       Connection delay :
(MXBean(com.bea:Name=source02,Type=JDBCDataSourceRuntime).ConnectionDelayTime)
count                                               21.0
mean                                                90.0
std                                                  0.0
min                                                 90.0
25%                                                 90.0
50%                                                 90.0
75%                                                 90.0
```

```
max                                                  90.0

[8 rows x 231 columns]
```

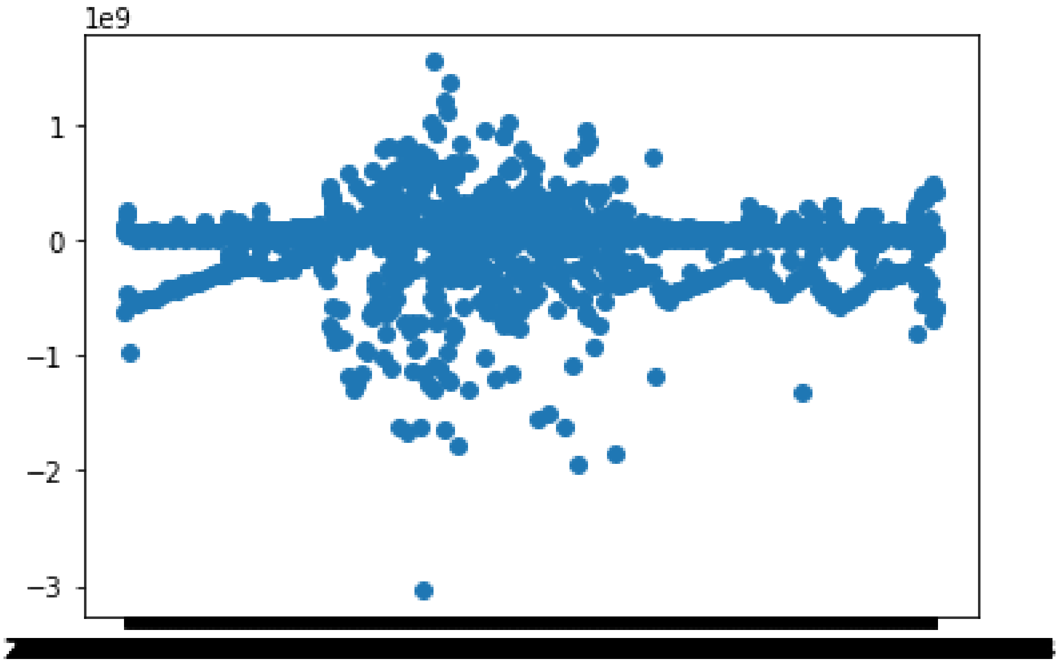
[74]:
```
#'||'.join('{}'.format(item) for item in df.columns)
```

[75]:
```
display(df.iloc[:,[2,10]])
df2 = df.iloc[:,[2,10]]
```

```
              timestamp  \
0      2015-02-02 21:45
1      2015-02-02 21:46
2      2015-02-02 21:47
3      2015-02-02 21:48
4      2015-02-02 21:49
...                 ...
40247  2015-03-02 20:49
40248  2015-03-02 20:50
40249  2015-03-02 20:51
40250  2015-03-02 20:52
40251  2015-03-02 20:53

       Heap usage activity : (d/dx (MXBean(java.lang:type=Memory).HeapMemoryUsage.used))
0                                           6.291546e+07
1                                          -6.303043e+08
2                                           1.255756e+08
3                                           1.232997e+08
4                                           2.080430e+08
...                                                  ...
40247                                      -2.024852e+09
40248                                       9.110411e+08
40249                                       7.840594e+08
40250                                      -2.055029e+09
40251                                       5.452400e+08

[40252 rows x 2 columns]
```

[76]:
```
fig = plt.figure()
ax1 = fig.add_subplot(111)
#ax1.scatter(otulierIndx[:,0], otulierIndx[:,1])
#ax1.scatter(normalIndx[:,0], normalIndx[:,1])
ax1.scatter(df2.iloc[0:2000,0], df2.iloc[0:2000,1])
plt.show
```

[76]:
```
<function matplotlib.pyplot.show(*args, **kw)>
```

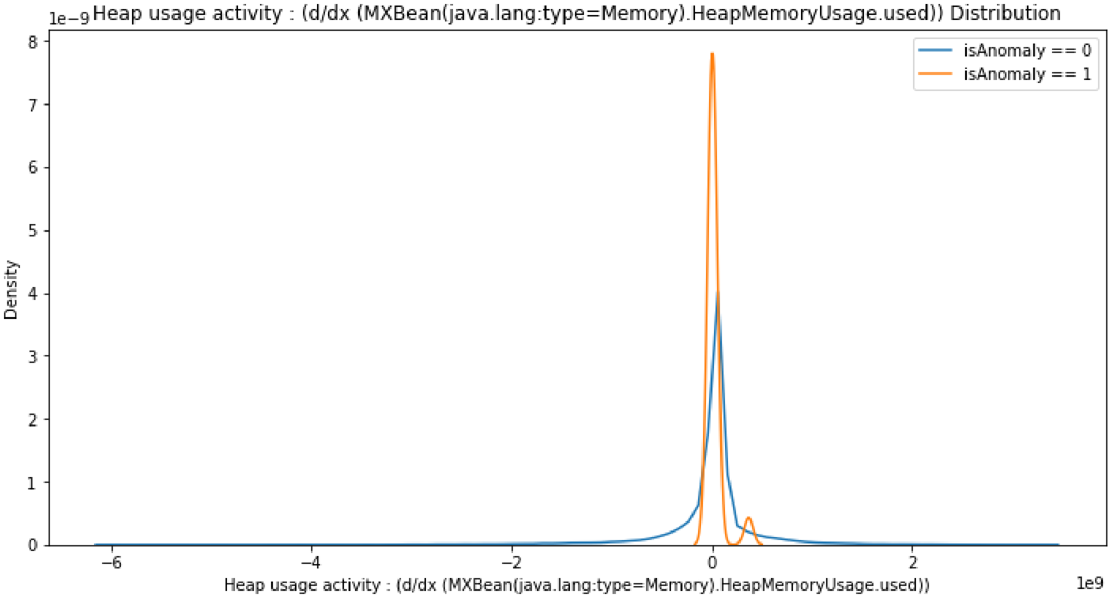


```
[77]: def kde_target(var_name, df):

          # Calculate the correlation coefficient between the new variable and the␣
       ↪target
          corr = df['isAnomaly'].corr(df[var_name])

          # Calculate medians for repaid vs not repaid
          avg_highr = df.loc[df['isAnomaly'] == 0, var_name].median()
          avg_lowr = df.loc[df['isAnomaly'] == 1, var_name].median()

          plt.figure(figsize = (12, 6))

          # Plot the distribution for target == 0 and target == 1
          sns.kdeplot(df.loc[df['isAnomaly'] == 0, var_name], label = 'isAnomaly ==␣
       ↪0')
          sns.kdeplot(df.loc[df['isAnomaly'] == 1, var_name], label = 'isAnomaly ==␣
       ↪1')

          # label the plot
          plt.xlabel(var_name); plt.ylabel('Density'); plt.title('%s Distribution' %␣
       ↪var_name)
          plt.legend();

          # print out the correlation
```

```
    print('The correlation between %s and the TARGET is %0.4f' % (var_name,␣
↪corr))
    # Print out average values
    print('Median value for request with high runtime value = %0.4f' %␣
↪avg_highr)
    print('Median value for request with low runtime value =     %0.4f' %␣
↪avg_lowr)
```

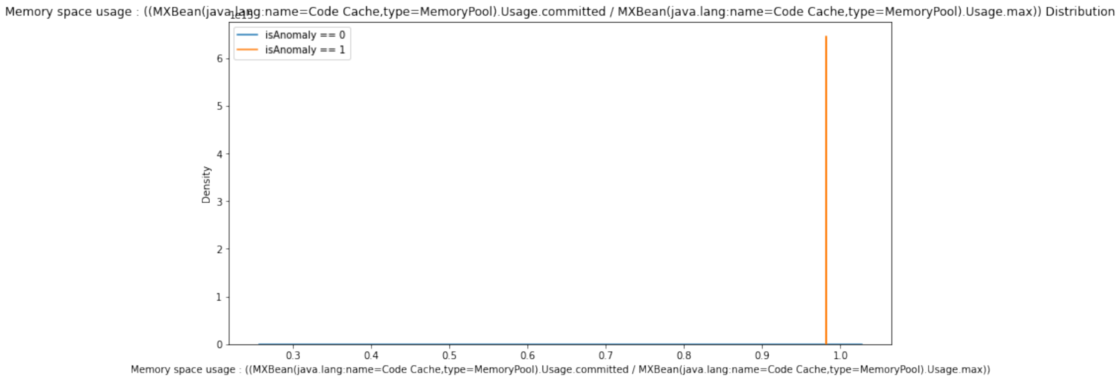
[78]:
```
kde_target('Heap usage activity : (d/dx (MXBean(java.lang:type=Memory).
↪HeapMemoryUsage.used))', df[['Heap usage activity : (d/dx (MXBean(java.lang:
↪type=Memory).HeapMemoryUsage.used))','isAnomaly']].dropna(),)
```

```
The correlation between Heap usage activity : (d/dx
(MXBean(java.lang:type=Memory).HeapMemoryUsage.used)) and the TARGET is 0.0013
Median value for request with high runtime value = 58136488.0000
Median value for request with low runtime value =     14758192.0000
```



[79]:
```
kde_target('Memory space usage : ((MXBean(java.lang:name=Code␣
↪Cache,type=MemoryPool).Usage.committed / MXBean(java.lang:name=Code␣
↪Cache,type=MemoryPool).Usage.max))', df[['Memory space usage : ((MXBean(java.
↪lang:name=Code Cache,type=MemoryPool).Usage.committed / MXBean(java.lang:
↪name=Code Cache,type=MemoryPool).Usage.max))','isAnomaly']].dropna(),)
```

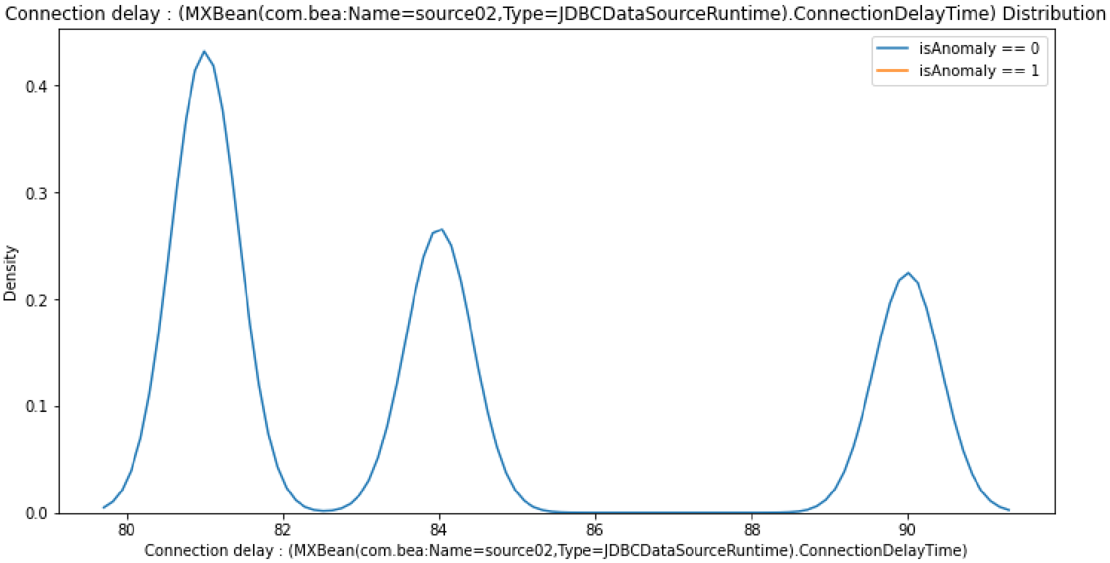
```
The correlation between Memory space usage : ((MXBean(java.lang:name=Code
Cache,type=MemoryPool).Usage.committed / MXBean(java.lang:name=Code
Cache,type=MemoryPool).Usage.max)) and the TARGET is 0.0128
Median value for request with high runtime value = 0.9701
Median value for request with low runtime value =     0.9818
```



[80]:
```
kde_target('Connection delay : (MXBean(com.bea:
 ↪Name=source02,Type=JDBCDataSourceRuntime).ConnectionDelayTime)',␣
 ↪df[['Connection delay : (MXBean(com.bea:
 ↪Name=source02,Type=JDBCDataSourceRuntime).
 ↪ConnectionDelayTime)','isAnomaly']].dropna(),)
```

```
The correlation between Connection delay :
(MXBean(com.bea:Name=source02,Type=JDBCDataSourceRuntime).ConnectionDelayTime)
and the TARGET is 0.0377
Median value for request with high runtime value = 84.0000
Median value for request with low runtime value =     90.0000
```

```
/home/swastik/.local/lib/python3.6/site-packages/seaborn/distributions.py:283:
UserWarning: Data must have variance to compute a kernel density estimate.
  warnings.warn(msg, UserWarning)
```


Connection delay : (MXBean(com.bea:Name=source02,Type=JDBCDataSourceRuntime).ConnectionDelayTime) Distribution
isAnomaly == 0
isAnomaly == 1
Density
Connection delay : (MXBean(com.bea:Name=source02,Type=JDBCDataSourceRuntime).ConnectionDelayTime)