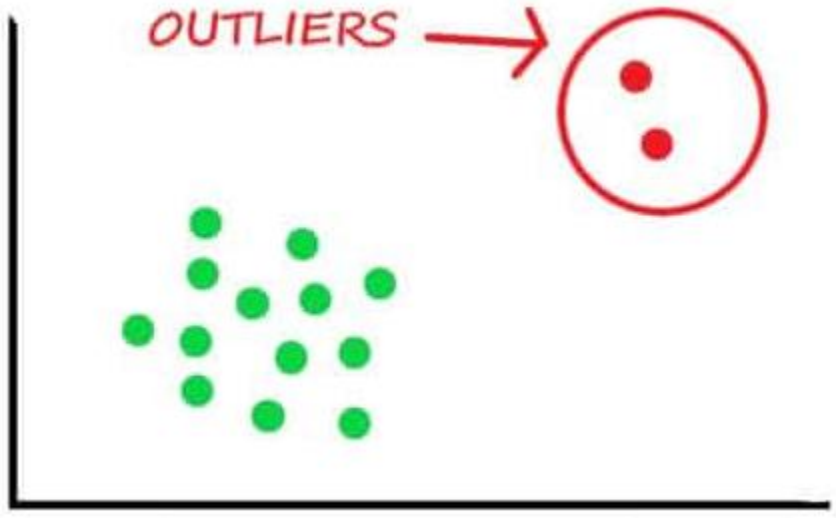
Maimoona Khilji
BS-DS
Semester 6

Lab Submission 06

Instructor: Basit Ali

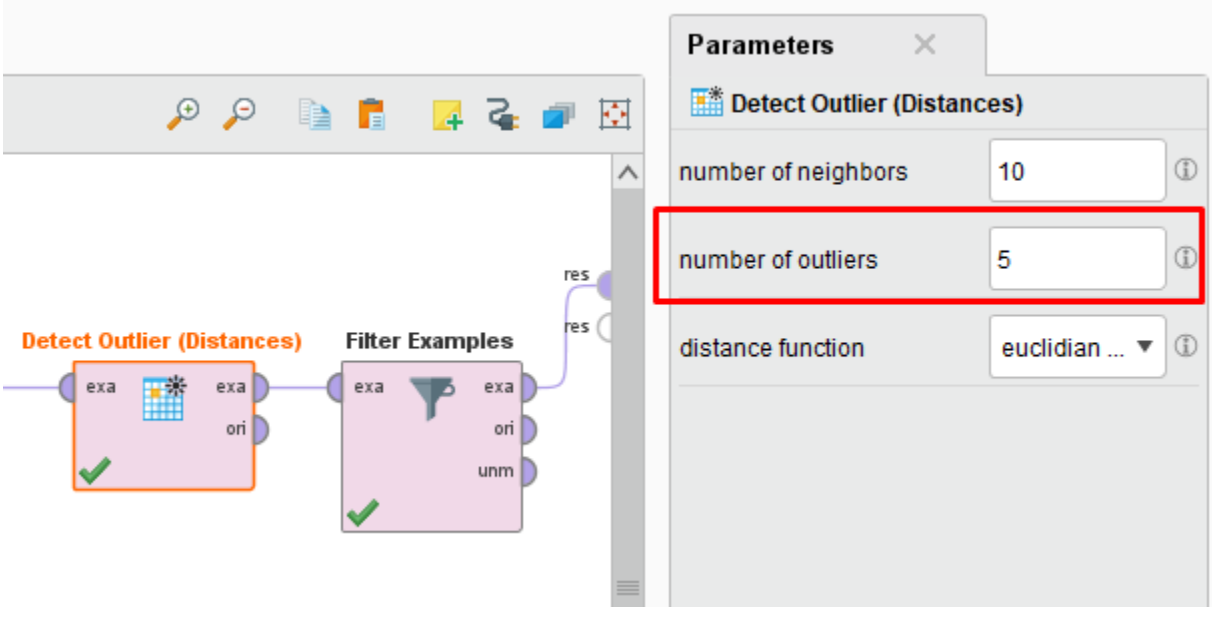
Q1. Write on note on outlier detection using Euclidean distance give example.

Detect Outlier is used to detect the outliers on the basis of finding the Euclidean distance between an object and its neighbors. It gives Boolean value in return. True represents outlier.



Q2. How would you change the process so it finds 20 outliers instead of 10?

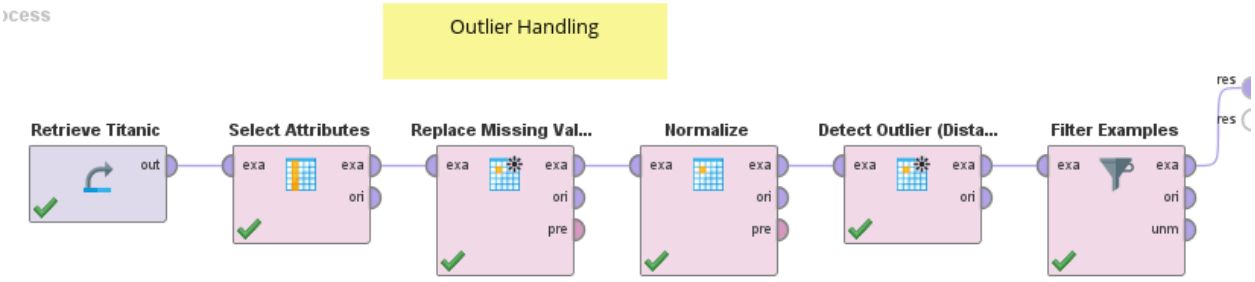
By changing the value of "number of outlier" in parameters section, we can change the process to find 20 outliers.



Maimoona Khilji
BS-DS
Semester 6

Q3. How can you change the process to only show outliers instead of removing them?

You can use **Filter example** to show only outliers instead of removing them.



Q4. Replace the outlier detection operator with Detect Outlier (LOF) What is the difference to before?

Detect Outlier (LOF) detect the outliers on the basis of local density in which locality is given by number of nearest neighbors whose distance is used to estimate the density. Then local density of an object is compared to its neighbor's local density. The points that have lower density than their neighbor is outlier.

The Detect Outlier (LOF) operator calculates an "Outlier Score" that helps you quickly summarize and find outliers in your dataset. When it finishes analyzing your dataset, it outputs a number: The larger the number in the Outlier column, the further away that data point is relative to your dataset.

Differences:

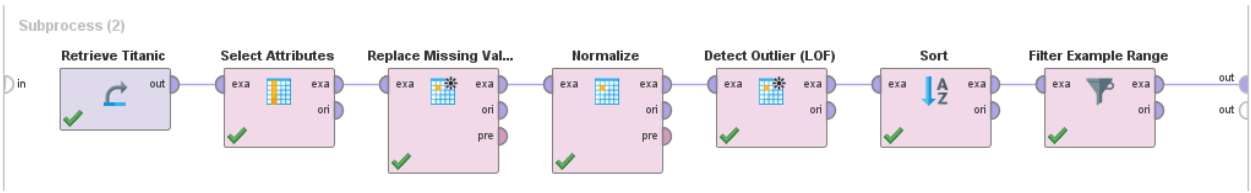
Detect Outlier (Distance Based)	Detect Outlier (LOF)
Find the outlier on the basis of distance.	Find the outlier on the basis of local density.
The **Detect outlier (Distance Based)** gives Boolean column.	**Detect Outlier (LOF)** gives numerical values for outlier column

Maimoona Khilji
BS-DS
Semester 6

Q5. How do you now need to change the filter to only keep the top outliers?

In order to show only top outliers, first we sort the **outlier** column in descending order. Then by using **Filter example range,** Select only 10 values.

These 10 values are top outliers.



Row No.	outlier	Age	No of Sibling...	No of Parent...	Passenger F...	Passenger ...	Sex	Port of Emb...	Survived
1	14323.348	0.000	0.481	-0.445	-0.519	Third	Male	Cherbourg	No
2	12144.054	0.785	-0.479	-0.445	-0.504	Third	Male	Cherbourg	No
3	10027.384	0.000	0.481	-0.445	-0.364	Third	Male	Cherbourg	No
4	9208.141	0.000	0.481	-0.445	-0.364	Third	Male	Cherbourg	No
5	8823.125	-0.224	0.481	-0.445	-0.364	Third	Male	Cherbourg	No
6	7730.389	-0.883	-0.479	-0.445	-0.504	Third	Male	Cherbourg	No
7	5934.298	0.000	-0.479	-0.445	-0.504	Third	Female	Cherbourg	Yes
8	5934.294	0.000	-0.479	-0.445	-0.504	Third	Female	Cherbourg	Yes
9	5648.025	0.009	-0.479	-0.445	-0.405	Second	Female	Queenstown	Yes
10	5647.787	0.000	-0.479	-0.445	-0.405	Second	Female	Queenstown	Yes