SPATIAL CLUSTER ANALYSIS

AND

COMPARISON OF THE ALGORITHMS

1. INTRODIUTION

a) Briefly description of the research

Clustering is dividing a dataset into groups called clusters without specifying the input data property requirements. Each cluster consists of objects similar to each other.

Cluster analysis is an unsupervised machine learning method for detecting patterns in data, it differs from the supervised machine learning method classification that results in labels.

The aim of the investigation is compare clustering algorithms and selecting the most appropriate one for working with the Cracow Crime data set.

During the research we will pay attention to:

- Complexity

- Execution Time

- Handling Noise

- Handling Complex Structures

- Number Of Clusters

- Dealing With Our Dasta Set

We will test 9 different clustering methods:

- K-Means
- HDBSCAN
- OPTICS
- BIRCH
- Spectral-clustering
- DBSCAN
- MeanShift
- AgglomerativeClustering (linkage=Ward)
- Gssian Mixture

Present the advantages and disadvantages of each of the above algorithms and when it is worth deciding on which one.

b) ESDA

In order to conduct a cluster analysis, we will use shp files that contain data on the city of Krakow and the location of crimes.

Shp format(shapefile) is a commonly used vector data interchange format developed by ESRI to represent point, line and area features along with their associated attribute data.

We will check whether the data is true or not made any mistakes when trying to measure and we will prepare them for use in algorithms.

ALGHORITM	COMPLEXITY	EXECUTION TIME	HANDLING NOISE	HANDLING COMPLEX STRUCTURES	NUMBER OF CLUSTERS	DASTA SET
KMeans	O(tkn)	0.9010	0	0	REQUIRED	CRACOW CRIMES
HDBSCAN	O(max{c^6 , c2^p , c2^l } c^10 N log Nα(N))	0.9101	1	1	NO REQUIRED	CRACOW CRIMES
OPTICS	O(n^2)	2.2502	1	1	NO REQUIRED	CRACOW CRIMES
BIRCH	Possible O(n)	1.1673	1	0.5	REQUIRED	CRACOW CRIMES
SPECTRAL_CLUSTERING	O(n^3)				REQUIRED	CRACOW CRIMES
DBSCAN	O (nlogn)	0.8249	1	1	NO REQUIRED	CRACOW CRIMES
MeanShift	O(n^2)	3.1125	0.5	0	NO REQUIRED	CRACOW CRIMES
AgglomerativeClustering (linkage=Ward)	DEPENDS OF HIERARHIAL ALGHORITM	1.0791	0	0	REQUIRED	CRACOW CRIMES
Gaussian Mixture	O(NKD3)	0.9085	0	0.5	REQUIRED	CRACOW CRIMES