Name: John Royce C. Punay   Course: PSMDSSC-101 - Data Mining

Student no: 2013048

1.) Visit the T.I.P. library and download a paper from the IEEE Digital Library. The paper should be related to supervised learning. Synthesize the paper using the Trend-Problem-Issue-Objectives-Contribution approach.

➔Trend

◆ This paper is all about the **unsupervised K-means clustering algorithm.** K-means is one of the simplest and popular unsupervised machine learning algorithms. The concept of unsupervised is that during the process, the system does not have concrete data sets in which the outcomes to most of the problem is unknown. In the simplest form, the machine learning algorithm is blinded in operation. A cluster refers to the collection of data sets consolidated together because of certain similarities and dissimilarities. In addition, clustering is the process of exploratory data analysis to get an initial understanding of the data sets. The task is identifying the same or subgroup, in other words, the objective is to find homogenous subgroups within the data points in each cluster.

➔Problem

◆ Kmeans algorithm is the popular model in machine learning. The usual behavior is partitioning data sets into initialized target number *k* but in general, the cluster number is unknown. Because the clustering itself is unsupervised, no truth is available to verify accurate results. The absence of truth complicates assessing quality. The second problem is clustering data is varying sizes and density.

➔Issues

◆ In general, the methods for partitioning is being represented by the finite cluster of prototypes associated with objective functions. It is essential to define the dissimilarity and distance between a point and a cluster prototype. According to this paper, k-means clustering is widely studied from different areas, but the actual issue is the algorithm is affected by the pre-defined *k.* Users need to specify a

range of clusters numbers in which the actual number of clusters is and then model selection.

➔Objectives

◆ The objective is to construct a learning procedure for the k-means clustering algorithm by automatically find the number of clusters without any initialization and parameter selection. Apply external and internal cluster validity indices. The external indices are used to evaluate clustering results by comparing cluster membership that is assigned by the clustering algorithm. Internal indices are used to evaluate the validity of structure by utilizing information about the data.

➔Contributions

◆ The contribution of X-means helps the clustering algorithm to make a local decision for clustering numbers without given a number of *k*. The usage of the entropy concept makes the clustering algorithm free of initializing the number instead, automatically find the number of clusters.

https://ieeexplore.ieee.org/document/9072123/