5

UNSUPERVISED LEARNING

Said Jadid Abdulkadir

5.1 INTRODUCTION

In this chapter a clear description of unsupervised machine learning is explained. Unsupervised learning is a form of self-organised learning that aids in discovering previously unrecognised patterns in a dataset without using labels. That implies that the learner learns with instances that are not labelled [37]. In fact, only the inputs are provided and no desired output, so in this type of the data will be grouped and then The model makes the comparison in order to predict the outcome [38]. Clustering and dimensionality reduction are common tasks in this learning style, where the learner must investigate essential patterns or correlations in the data to understand relations instead of rules [39]. The critical discrepancy between supervised and unsupervised learning is that, there is no comprehensive and clean labelled dataset in unsupervised learning [40]. Figure 5.1 illustrates the common types of unsupervised learning approach works. In the following sections we will explain two types of unsupervised learning, dimensionality reduction and K-means clustering.

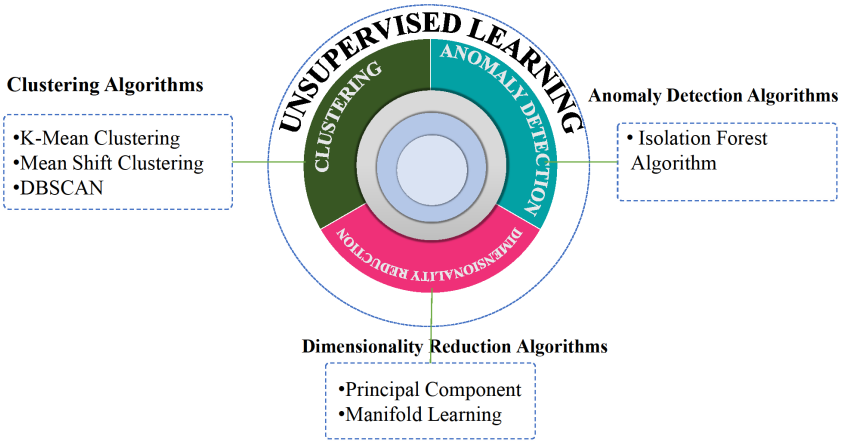

Figure 5.1 Unsupervised learning

Unsupervised learning is a form of self-organised learning in which previously unknown correlations in a data set are discovered through pre-existing labels [41]-[42]. This means that the learner is learning from unlabeled examples. In practice, only the inputs are given, with no desired output, so in this case, the data will be clustered, and the model will perform a comparison to guess the output. Clustering, Anomaly Detection, and Dimensionality Reduction are typical activities in this type of learning, in which the learner must analyse important patterns or correlations in data to consider relations rather than rules. Figure 5.2 demonstrates how unsupervised learning technique works.

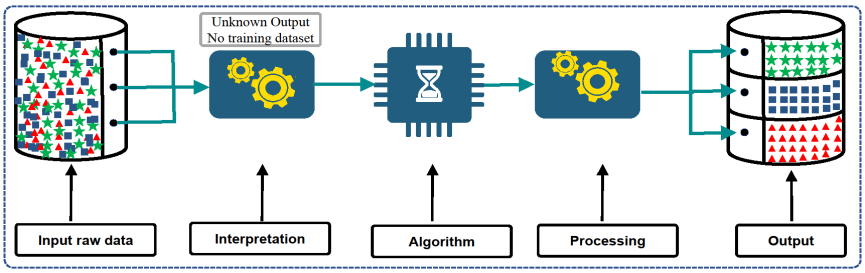

Figure 5.2 Unsupervised learning process

5.2 CLUSTERING MODELS

Clustering is an unsupervised learning strategy that is widely utilised in various disciplines for data analysis. When applying a clustering algorithm to input data, it can observe what categories the data points fit into, providing some useful insights [43]-[44]. The following subsections go through three common clustering techniques in depth.

5.2.1 K-Mean Clustering

The K-Mean algorithm is probably the most commonly unsupervised learning technique used in clustering problems. Most of interdiction classes in data science and machine learning adapt it to teach beginners [45]. It divides the data set into k-clusters to keep each cluster as compact as feasible while separating the k-clusters as much as possible. The distances between the cluster objects and the cluster centre are used to define the related cost function. The goal is to identify k partitions that reduce the cost function to its smallest value.

The K-Means algorithm initiates and randomly originates the groups/classes that can be used in different perspective analyses to their respective centre points. It identifies distinct categories to find the number of classes that can be used. Each data point vector is the same length as the vectors of the centre points, and the groups are classified based on the closest point to the centre of each group by computing the distance between each point. The classified points are used to recompute the group centres using the mean of all vectors in