Chapter 2
A Study on the Clustering Techniques and Performance Evaluation

F. Mohamed Ilyas
https://orcid.org/0009-0002-8954-2449
The New College, India

S. Silvia Priscila
https://orcid.org/0000-0002-6040-3149
Bharath Institute of Higher Education and Research, India

V. Poornima
Vels Institute of Science, Technology, and Advanced Studies, India

Prasanna Ranjith Christodoss
Messiah University, USA

D. Kerana Hanirex
Bharath Institute of Higher Education and Research, India

C. Elayaraja
Dhaanish Ahmed College of Engineering, India

M. Sakthivanitha
https://orcid.org/0009-0001-6795-6555
Vels Institute of Science, Technology, and Advanced Studies, India

A. T. Ashmi Christus
Dhaanish Ahmed College of Engineering, India

ABSTRACT

People prefer approximations based on resemblance, particularly when dealing with commercial figures. With data, the procedure is even more exact, known as Clustering. Clustering is a method that has numerous applications. Numerous fields, including pattern identification, image analysis, consumer statistical analysis, segmentation of markets, social network analysis, and more, use clustering, an unsupervised machine learning technique. It can effectively address many problems and goals, from the most basic to the most complicated. Cluster analysis is an effective approach for detecting underlying structures and patterns in data from diverse areas. It divides

DOI: 10.4018/979-8-3373-5203-9.ch002

the data into clusters based on their similarities or differences. Researchers and analysts can get significant insights and make data-based decisions by combining comparable observations. Further, Clustering on the original information provided us with extremely valuable and unique explanations for each of the individual segments.

INTRODUCTION

Clustering is the process of grouping a population or set of data points into groups so that data points in the same category are more similar than those in other groups. In other words, the goal is to separate groups with similar characteristics and assign them to clusters. Analyzing data is identifying meaningful data in unstructured data. Machine learning (ML) is one of the most extensively utilized data analysis methodologies (Ghosh and Kumar, 2013). It is a framework or system that can derive information from unorganized information and make judgments without requiring human involvement. The most common ML strategies used for data analysis are classification and Clustering. It incorporates numerous algorithms to categorize items into classes based on their properties (Tareq et al., 2024). Clustering analysis, as a subfield of data mining (DM) and machine learning (ML), has significant consequences for examining and identifying data structures, processing data, and choosing features. Similarly, Clustering, as an unsupervised learning technique, seeks to split data without previous instruction. (Bhardwaj et al., 2024). Its goal is to make certain that objects in the same cluster are as similar as feasible and that items in other groups are as distinct as possible. Clustering analysis has several uses in today's information age, including picture segmentation, medical evaluation, and social media analysis.

Techniques for Clustering employ the fundamental framework of the data to develop rules for organizing data with similar characteristics, dividing the data set based on group characteristics without prior knowledge of the information. In a perfect example, every group of clusters has a data instance that is more related to one another than instances belonging to other clusters. (Troiano et al., 2024). The overview of Clustering is presented in Figure 1. The raw data is clustered into f=different groups of data