**Graph-based Clustering and K-mean Clustering**

ABSTRACT : This research focuses on exploring and comparing two distinct approaches to clustering analysis—Graph-Based Clustering and K-Means Clustering—applied to a specific dataset. The dataset chosen for this study is related to Titanic survival predictions, a classic dataset in machine learning. Graph-Based Clustering methods, including techniques such as k-Nearest Neighbours (KNN) and Epsilon Neighbourhood Graph, leverage the inherent relationships between data points to form clusters. Additionally, minimum spanning trees are employed through algorithms like Prim's and Kruskal's to visualize and analyse the resulting clusters. On the other hand, K-Means Clustering, a centroid-based algorithm, partitions the data into clusters based on the minimization of within-cluster variance. The research involves implementing these clustering techniques, evaluating their performance, and comparing the obtained clusters. The study also considers the application of visualization tools to enhance the interpretation of the clustering results. Through this investigation, insights into the strengths and limitations of each clustering approach are gained, contributing to a comprehensive understanding of their applicability in real-world scenarios.

INDEX TERM : Graph-Based Clustering, K-Means Clustering, Titanic Dataset, k-Nearest Neighbors (KNN), Epsilon Neighborhood Graph, Minimum Spanning Tree, Prim's Algorithm, Kruskal's Algorithm, Centroid-Based Clustering, Data Visualization, Machine Learning, Data Clustering Unsupervised Learning, Feature Engineering, Dimensionality Reduction, Predictive Modeling, Titanic Survival Prediction, Data Analysis, Evaluation Metrics (e.g., Silhouette Score, Confusion Matrix)

|  |  |
| --- | --- |
|  |  |