Data Science Project

Customer
Segmentation

<u>Abstract</u>

Today's business run on the basis of such innovation having ability to enthral the customers with the products, but with such a large raft of products leave the customers confounded, what to buy and what to not and also the companies are nonplussed about what section of customers to target to sell their products. This is where machine learning comes into play, various algorithms are applied for unravelling the hidden patterns in the data for better decision making for the future. This elude concept of which segment to target is made unequivocal by applying segmentation. The process of segmenting the customers with similar behaviour's into the same segment and with different patterns into different segments is called customer segmentation. In this paper, 3 different clustering algorithms (k-Means, Agglomerative, and Meanshift) are been implemented to segment the customers and finally compare the results of clusters obtained from the algorithms. By applying clustering, 5 segments of cluster have been formed labelled as Careless, Careful, Standard,Target and Sensible customers. However, two new clusters emerged on applying mean shift clustering labelled as High buyers and frequent visitors and High buyers and occasional visitors.

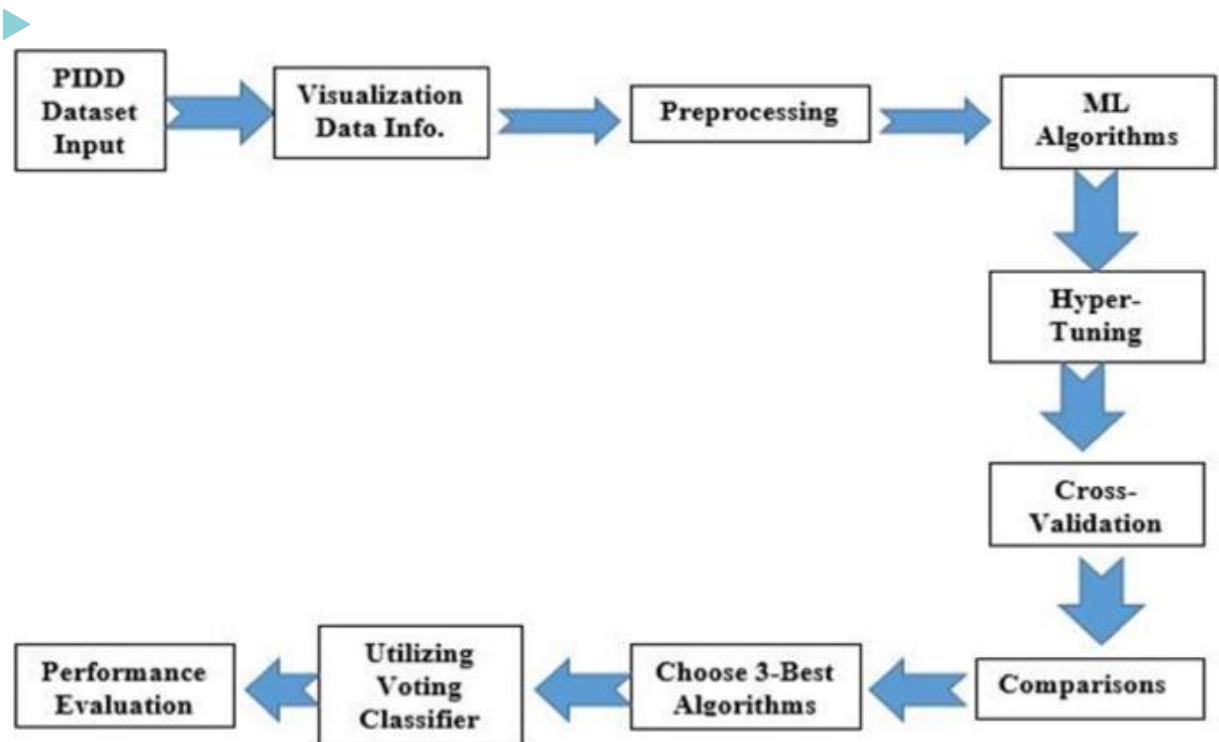# Contents:

<u>Introduction</u>

In Unsupervised Machine Learning, we do not need to supervise the model. Such a method deals with unlabelled data. Unsupervised machine learning helps us find hidden and unknown patterns in data.

Often it easier to get unlabelled data as compared to labelled data, and in such cases, we can use unsupervised machine learning to work on the data. Data, which needs categorization can be categorized with the help of unsupervised machine learning.

Clustering is a type of unsupervised machine learning in which the algorithm processes our data and divided them into "clusters".

Clustering algorithms try to find natural clusters in data, the various aspects of how the algorithms to cluster data can be tuned and modified. Clustering is based on the principle that items within the same cluster must be similar to each other. The data is grouped in such a way that related elements are close to each other.

Methodology

▶

| PIDD Dataset Input | → | Visualization Data Info. | → | Preprocessing | → | ML Algorithms |

(ML Algorithms → Hyper-Tuning → Cross-Validation → Comparisons → Choose 3-Best Algorithms → Utilizing Voting Classifier → Performance Evaluation)

▶ The Proposed method use e K-means clustering algorithm for clustering of customers using trained data.

Implementation

Mall Customer data is an interesting dataset that has hypothetical customer data. It puts you in the shoes of the owner of a supermarket. You have customer data, and on this basis of the data, you have to divide the customers into various groups.

▶ The data includes the following features:
▶ 1. Customer ID
▶ 2. Customer Gender
▶ 3. Customer Age
▶ 4. Annual Income of the customer (in Thousand Dollars)
▶ 5. Spending score of the customer (based on customer behaviour and spending nature)

Conclusion

K means clustering is one of the most popular clustering algorithms and usually the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset. The goal of K means is to group data points into distinct non-overlapping subgroups. One of the major application of K means clustering is segmentation of customers to get a better understanding of them which in turn could be used to increase the revenue of the company.