Customer Clustering with K-Means

Selvaraj.X(2577127)

07-NOV-2023

Introduction

- E-commerce companies have a large amount of customer data, which can be used to understand their customers better and improve their marketing and recommendations.
- Customer clustering is a technique that can be used to group customers into distinct clusters based on their shopping behavior and preferences.
- K-Means clustering is a simple and effective algorithm for customer clustering.

Importance of customer clustering

- Customer clustering helps e-commerce companies to understand their customers better and their shopping behavior.
- E-commerce companies can use customer clustering to target their marketing campaigns and personalized recommendations more effectively.
- For example, e-commerce companies can send targeted email campaigns to customers in different clusters, or recommend products that are relevant to their interests.

Data Preprocessing

- Scaling: Numerical features should be scaled to the same range so that they have equal importance in the clustering process.
- Handling missing values: Missing values can be handled by imputing them with the mean, median, or mode of the feature.
- Encoding categorical variables: Categorical variables need to be converted into numerical values before using the K-Means clustering algorithm. This can be done using a technique called label encoding.

Implementing K-Means

```
# K-means model

clust=KMeans(n_clusters=3)
clust.fit(data)
```

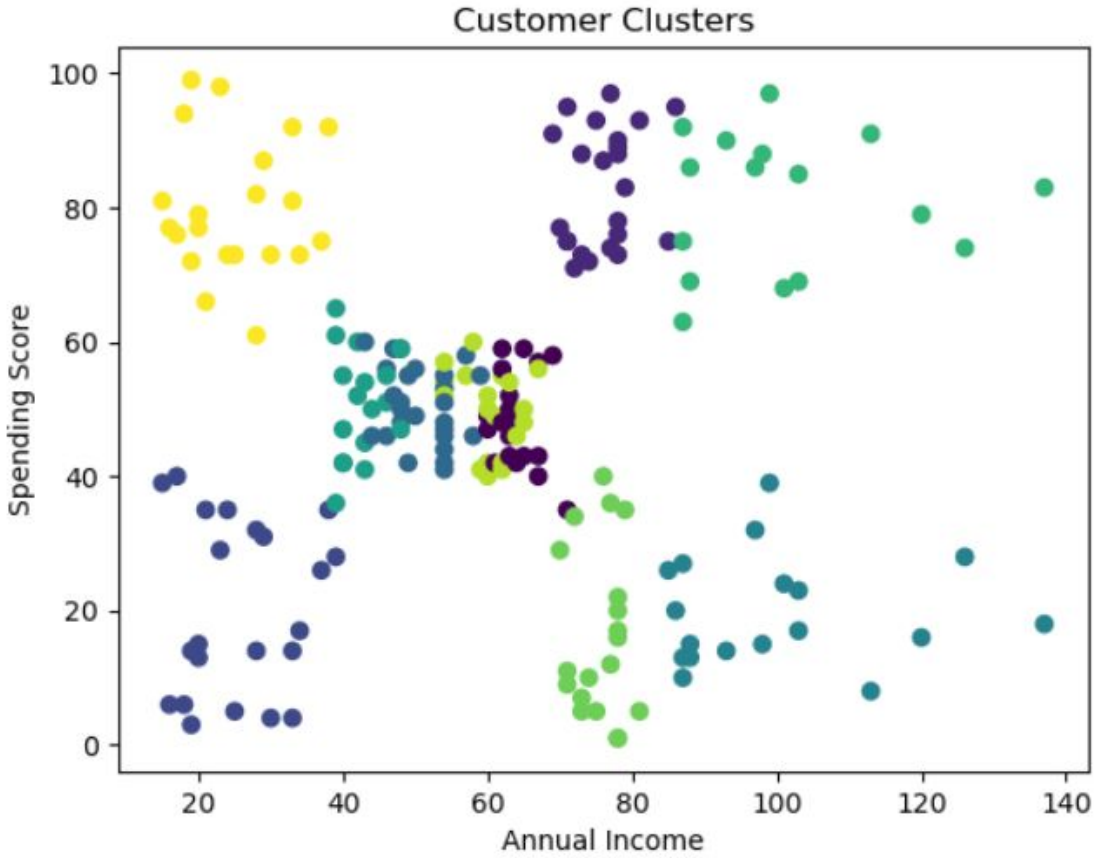
```
k=list(range(2,11))
y=[]
for i in k:
    km=KMeans(n_clusters=i,random_state=1,init='k-means++')
    km.fit(data)
    labels=km.labels_
    y.append(km.inertia_)
plt.plot(k,y)
plt.xlabel('No.of Cluesters')
plt.ylabel('Error')
```

Model Training and Evaluate Performance

```
#Performance measure
from sklearn.metrics import silhouette_score
labels=clust.labels_
print(labels,len(labels))
print('Silhouette Score:\t',silhouette_score(data,labels))
```

```
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0] 200
Silhouette Score:        0.3766801400675223
```

Visualize the customer clusters



Hyperparameter Tuning

```
# Hyperparameter Tuning

clust=KMeans(n_clusters=2)
clust.fit(data)

labels=clust.labels_
print(labels,len(labels))
print('Silhouette Score:\t',silhouette_score(data,labels))
```

```
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0] 200
Silhouette Score:        0.4793049017423563
```

★ **k = 2 gives high accuracy, then other k values. K = 2 is give optimum results**

Real-World Application

- Targeted marketing: E-commerce companies can use customer clustering to target their marketing campaigns more effectively.
- Personalized recommendations: E-commerce companies can use customer clustering to generate more personalized recommendations for their customers.
- Customer retention: E-commerce companies can use customer clustering to identify and retain their most valuable customers.
- Fraud detection: E-commerce companies can use customer clustering to detect fraudulent transactions.

Potential limitations of the K-Means clustering algorithm

- Sensitivity to the initial cluster centroids: The K-Means clustering algorithm randomly initializes the cluster centroids at the beginning of the algorithm. This means that the final clustering results can vary depending on these initial positions. Different initializations can lead to different local optima, resulting in different clustering outcomes.
- Assumption of spherical and equal-sized clusters: The K-Means clustering algorithm assumes that the clusters are spherical and equal-sized. This may not be true for all real-world datasets, and can lead to inaccurate clustering results.
- Inability to handle outliers: Outliers can distort the cluster shapes and sizes, and can lead to inaccurate clustering results.
- Inability to handle categorical data: The K-Means clustering algorithm is designed to cluster numerical data. Categorical data needs to be converted into numerical values before using the K-Means clustering algorithm. This conversion can be challenging, and can lead to inaccurate clustering results.

Conclusion

- Customer clustering is a powerful tool that can be used by e-commerce companies to understand their customers better and improve their marketing and recommendations.
- The K-Means clustering algorithm is a simple and effective algorithm for customer clustering.
- However, it is important to be aware of the limitations of the K-Means clustering algorithm and to take steps to mitigate them.