Challenge 19: Unsupervised Learning

You will find a jupyter notebook to create a machine learning model that groups cryptocurrencies:

1. I used the sklearn library in Python to load the unclustered cryptocurrency data set and explore the data. There are 7 columns.

2. While preprocessing or preparing the data, there are no nulls to eliminate, but I used StandardScaler to scale the data, placing all features in the same range.

3. Imported the K-Means model from the sklearn library. Using the elbow method ran the K-Means and the plot suggests that k=4 is the optimal number of clusters for the dataset. As we added more clusters k= 3, 4, we diminished the improvement of lowering the total distance or inertia values; the distance diminished at a slower pace. The silhouette score of .31 at k=4 was not close to 1. This made k values not optimal for predicting 24-hour or 7-day price changes affecting cryptocurrencies.

4. Running PCA on scaled data to reduce dimensionality. I used the Principal Component Analysis (PCA), reducing the features from 7 to 3. Then, the top 2 components were plotted, which suggests that k=4 is the best value for k.

5. Used hvplot for data visualization. By comparing the cluster cryptocurrencies and PCA methods to find the best k value, the PCA helped understand the effect of dimensionality reduction on the clustering process. The clusters did not overlap with K Means; however, after the dimensionality reduction with PCA, there was a signal of more well-defined clusters.

Source

Xpert. (2024). AI Learning Assistant. EdX.