

## **Apply clustering and visualize using PCA**

**Clustering** is an **unsupervised learning technique** used to group similar data points based on shared features without predefined labels. In this assignment, **K-Means clustering** was applied to group tweets with similar text patterns. The algorithm divides the dataset into clusters by minimizing the distance between data points and their cluster centers.

Since text data is high-dimensional, **Principal Component Analysis (PCA)** was used to reduce the features to **two dimensions** for visualization. PCA helps capture the most important variance in the data, allowing the clusters to be displayed clearly in a 2D plot.

By applying clustering, we can identify underlying patterns or sentiments in the dataset without labeled data. This unsupervised analysis adds value to the project by helping understand the **natural groupings and trends** in the text.