Name: Minahil Irfan

Assignment 8: Unsupervised Learning – Clustering and PCA Visualization

Overview:
This assignment focused on applying unsupervised learning techniques to the Twitter sentiment dataset. K-Means clustering was used to group similar tweets, and PCA was applied to visualize the high-dimensional data in 2D.

Key Concepts:

1. Clustering:

- Unsupervised learning technique used to group similar data points.
- K-Means algorithm divides the dataset into clusters by minimizing the distance between points and their cluster centers.
- Helps identify natural patterns or trends without using labels.

2. Principal Component Analysis (PCA):

- Dimensionality reduction technique.
- Reduces high-dimensional data to two dimensions for visualization.
- Captures the most important variance in the data, making clusters easier to observe.

Implementation Steps:

1. **Load Dataset:**
    - Loaded twitter_sentiment_cleaned.csv into Pandas.
    - Dropped missing values in the text column.
2. **Feature Engineering:**
    - Converted tweets into numeric features using CountVectorizer (max 1000 features).
3. **Clustering:**
    - Applied K-Means clustering with 2 clusters to group similar tweets.
    - Added cluster labels to the dataset for analysis.
4. **Dimensionality Reduction & Visualization:**
    - Applied PCA to reduce the high-dimensional feature space to 2 components.
    - Plotted the clusters in a 2D scatter plot using Matplotlib.
5. **Analysis:**
    - Examined a few samples from each cluster to understand underlying patterns.

Insights & Conclusion:

- K-Means clustering revealed natural groupings of tweets based on text similarity.
- PCA visualization helped in understanding cluster separation and structure in 2D.
- Unsupervised analysis complements supervised methods, providing additional understanding of patterns in the dataset.

Project Milestone:

Added unsupervised learning analysis to the project, enhancing exploratory insights.

Github: DataScience-AI/assigmet8 at main · MinahilIrfan98/DataScience-AI