Results

SBERT

Atis_data

```
mean_RI_score 0.7275063385913116
mean_ARI_score 0.10960661552739129
total time: 36.0 sec

Process finished with exit code 0
```

News_mix_data

```
mean_RI_score 0.8352290424113097
mean_ARI_score 0.6307049856024117
total time: 12.0 sec

Process finished with exit code 0
```

TFIDF

Atis_data

```
mean_RI_score 0.7277449636052566
mean_ARI_score 0.10567782306852973
total time: 31.0 sec

Process finished with exit code 0
```

News_mix_data

```
mean_RI_score 0.5522661047390415
mean_ARI_score 0.05899913446829981
total time: 14.0 sec

Process finished with exit code 0
```

Performance difference

The dataset with the bigger performance gap between TF-IDF and SBERT is **news_mix_data**.

On the news dataset, SBERT clearly outperforms TF-IDF because news articles are longer and more diverse, and texts about the same topic often use different words. SBERT captures the semantic meaning of the text, while TF-IDF mainly relies on word overlap, which leads to poor clustering.

Clusters inspection

Example 1

- Text: Rachel Dolezal Faces Felony Charges For Welfare Fraud
- True Category: CRIME
- TFIDF: Clustered incorrectly
- SBERT: Clustered correctly

TF-IDF mainly relies on surface keywords such as *“welfare”* or *“fraud”*, which can also appear in political or social contexts, causing confusion. SBERT captures the overall meaning of the sentence and understands that it describes a criminal charge, so it places the text in the correct crime-related cluster.

Example 2

- Text: Hospice Overdosed Patients To 'Hasten Their Deaths,' Former Health Care Executive Admits
- True Category: CRIME
- TFIDF: Clustered incorrectly
- SBERT: Clustered correctly

TF-IDF is influenced by medical terms like *“hospice”* and *“health care”*, which can lead it to cluster the text with health-related articles. SBERT understands the semantic context of wrongdoing and admission of guilt, allowing it to correctly associate the text with criminal activity.