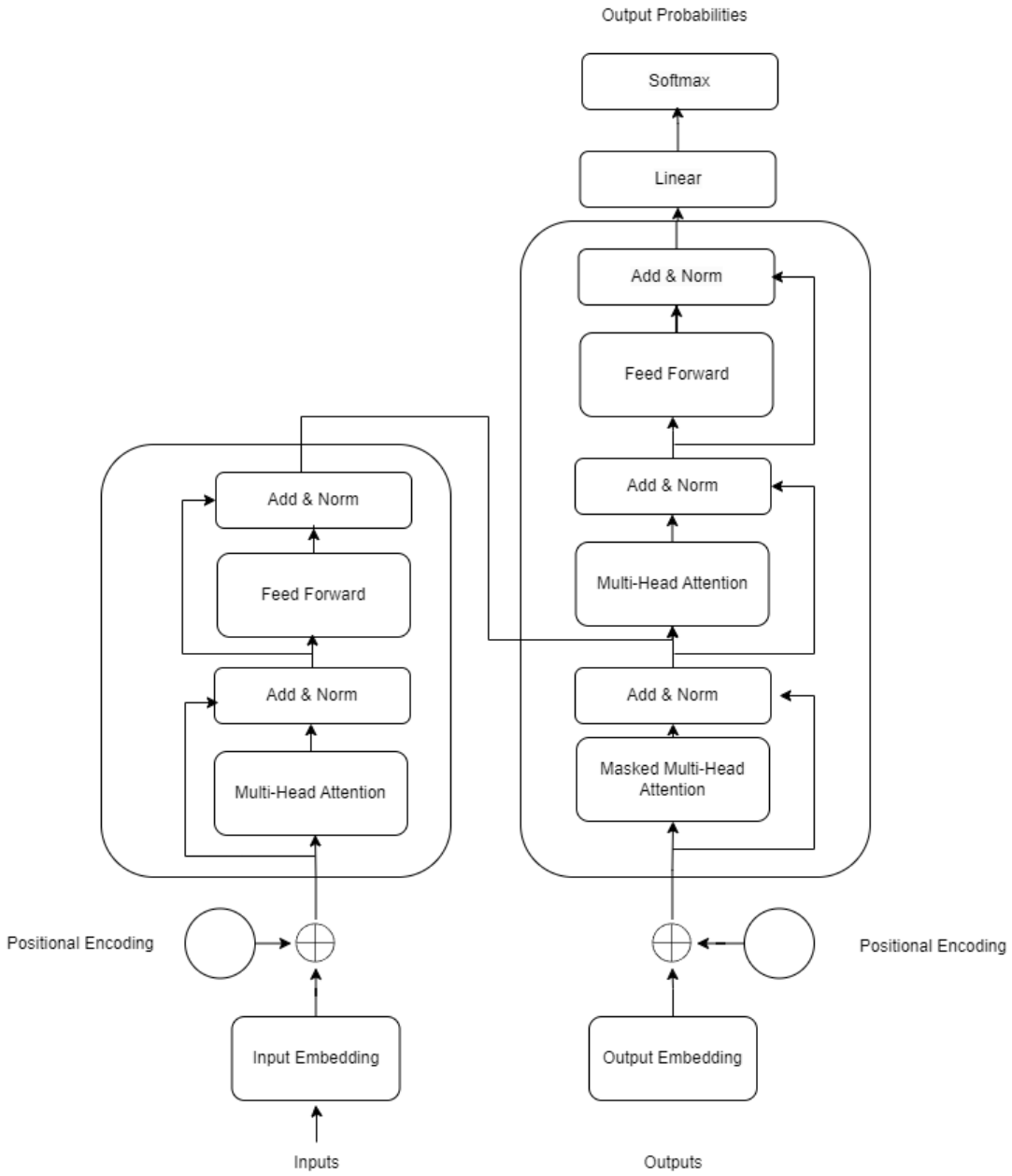
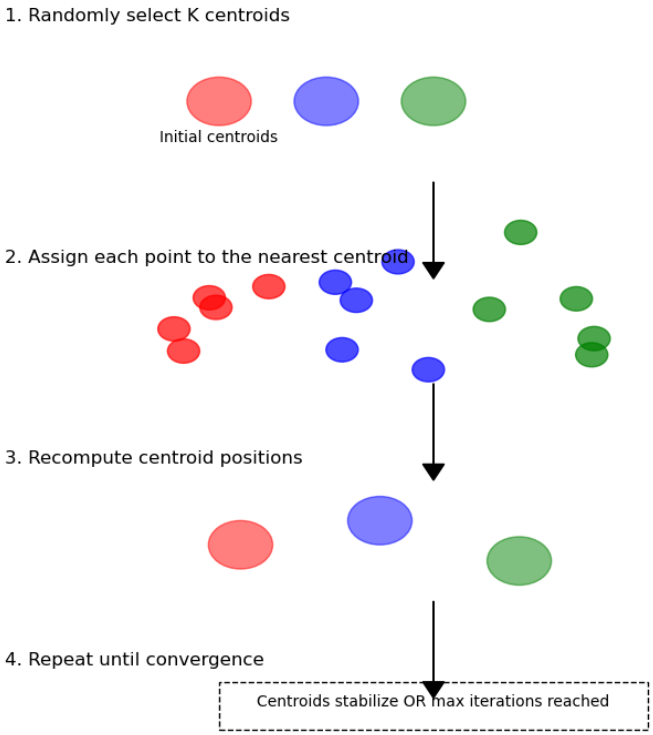
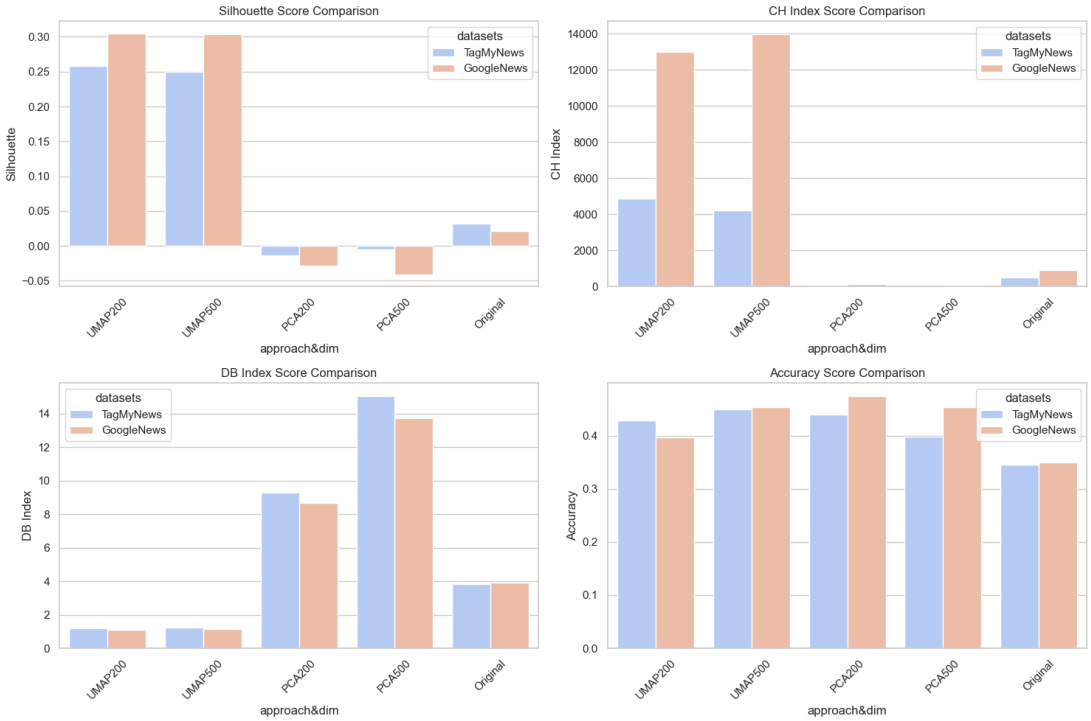
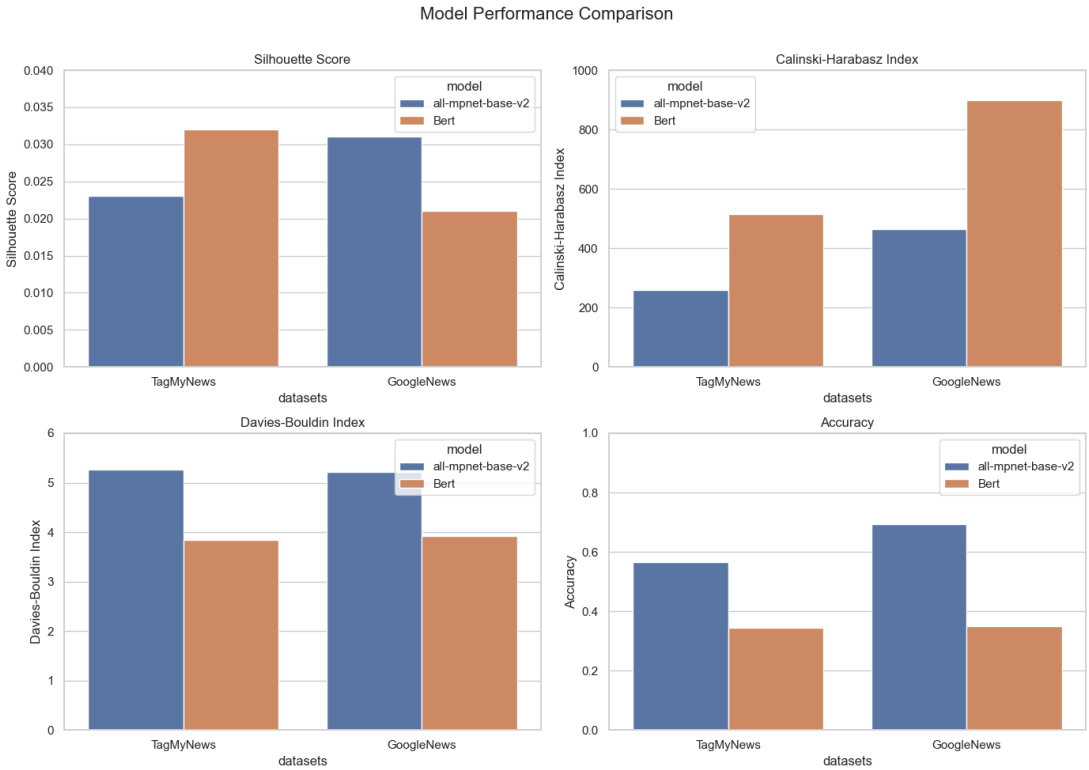
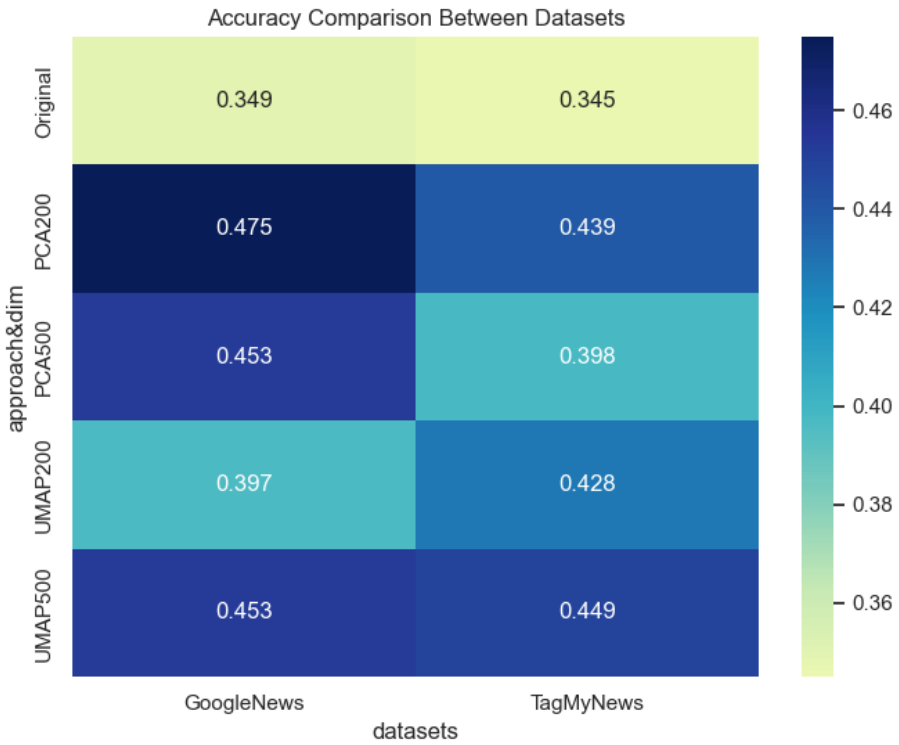

Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs

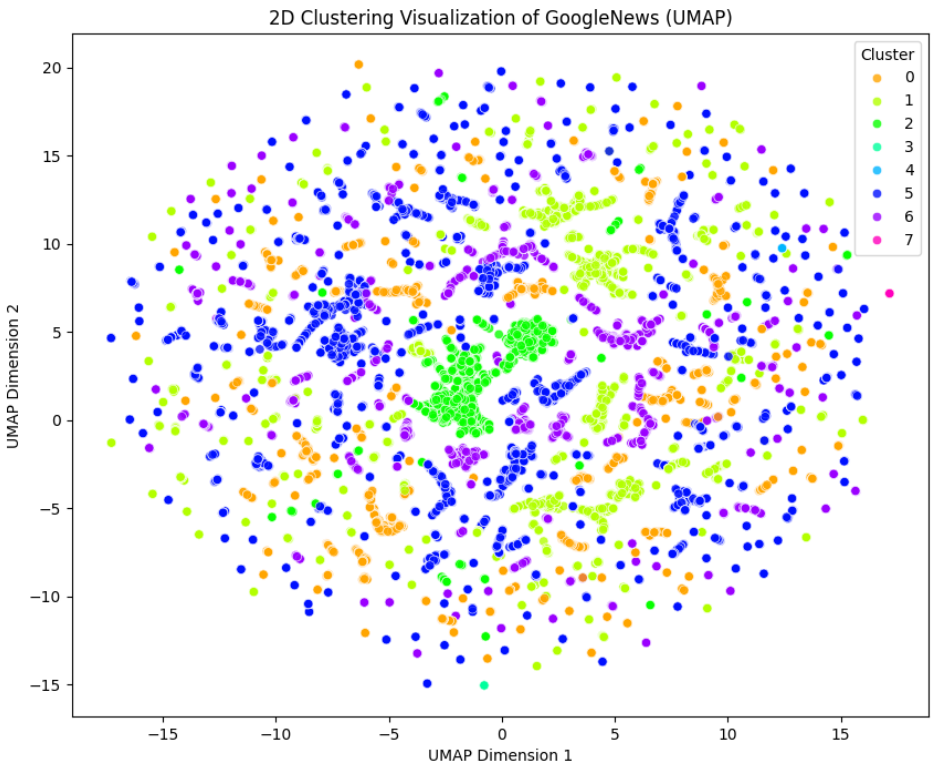
K-Means Clustering Algorithm Flowchart

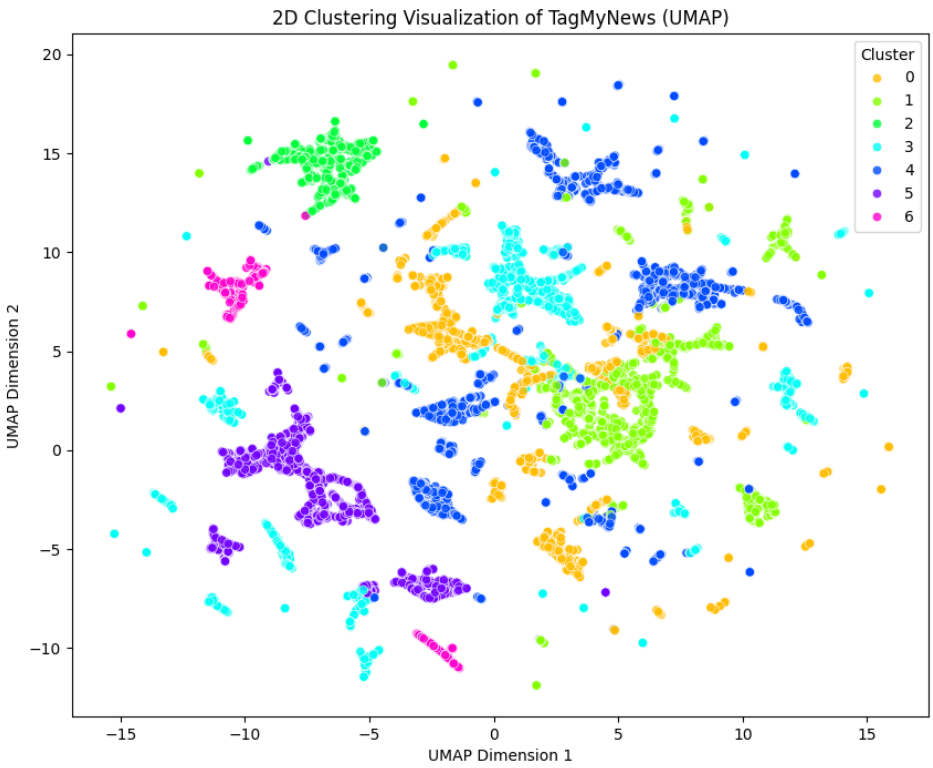



Silhouette Score Comparison
Silhouette
approach&dim
datasets
TagMyNews
GoogleNews
UMAP200
UMAP500
PCA200
PCA500
Original
CH Index Score Comparison
CH Index
DB Index Score Comparison
DB Index
Accuracy Score Comparison
Accuracy


Model Performance Comparison
Silhouette Score
Silhouette Score
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
model
all-mpnet-base-v2
Bert
TagMyNews
GoogleNews
datasets
Calinski-Harabasz Index
Calinski-Harabasz Index
1000
800
600
400
200
0
model
all-mpnet-base-v2
Bert
TagMyNews
GoogleNews
datasets
Davies-Bouldin Index
Davies-Bouldin Index
6
5
4
3
2
1
0
model
all-mpnet-base-v2
Bert
TagMyNews
GoogleNews
datasets
Accuracy
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
model
all-mpnet-base-v2
Bert
TagMyNews
GoogleNews
datasets


Accuracy Comparison Between Datasets
Original
PCA200
PCA500
UMAP200
UMAP500
approach&dim
0.349
0.345
0.475
0.439
0.453
0.398
0.397
0.428
0.453
0.449
GoogleNews
TagMyNews
datasets
0.46
0.44
0.42
0.40
0.38
0.36


2D Clustering Visualization of GoogleNews (UMAP)
Cluster
0
1
2
3
4
5
6
7
UMAP Dimension 2
UMAP Dimension 1
20
15
10
5
0
−5
−10
−15
−15
−10
−5
0
5
10
15


2D Clustering Visualization of TagMyNews (UMAP)
Cluster
0
1
2
3
4
5
6
UMAP Dimension 2
UMAP Dimension 1
20
15
10
5
0
−5
−10
−15
−10
−5
0
5
10
15